

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2004, 02:32:16 ; Search time 11655 Seconds
(without alignments)
11610.206 Million cell updates/sec

Title: US-10-088-384A-27
Perfect score: 3122
Sequence: 1 ttcataatgacgtcttc.....atgagatagccttgraagca 3122

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3122	100.0	3122	6	AX099282	Sequence
2	3122	100.0	81414	8	AB024037	Arabidops
3	3122	100.0	110157	8	AF058825	Arabidops
4	2029	65.0	74475	8	T7M24	Arabidops
5	2029	65.0	196296	8	ATCHRIV10	Arabidops
6	1973	63.2	68352	8	T15F17	Arabidops
7	1967.2	63.0	116966	8	AB046433	Arabidops
8	1937.2	62.0	81902	8	T17A2	Arabidops
9	1937.2	62.0	195837	8	ATCHRIV21	Arabidops
10	1925.4	61.7	123460	8	AB073166	Arabidops
11	1911.6	61.2	14879	8	AC084241	Arabidops
12	1911.6	61.2	117055	8	AC073433	Arabidops
13	1910.6	61.2	69752	8	T5H22	Arabidops
14	1910.6	61.2	159629	8	ATCHRIV9	Arabidops
15	1905.4	61.0	98536	8	AC074111	Arabidops
16	1904.4	61.0	96547	8	AB073154	Arabidops
17	1904.4	61.0	104628	8	AB073153	Arabidops
18	1904.4	61.0	137179	2	AC109923	Arabidops
19	1904	61.0	132522	8	AB062089	Arabidops
20	1900.2	60.9	102331	8	AC007120	Arabidops
21	1899.8	60.9	77287	8	AB026642	Arabidops
22	1897.6	60.8	114656	8	ATT15B3	Arabidops
23	1890	60.5	96684	8	AB046440	Arabidops
24	1890	60.5	116461	8	AC051625	Genomic S
25	1890	60.5	133529	8	AC069553	Genomic S
26	1880	60.2	118662	2	AC087569	Arabidops
27	1874.2	60.0	81902	8	T17A2	Arabidops
28	1874.2	60.0	110826	8	AC007125	Arabidops
29	1873.2	60.0	90840	8	AB073158	Arabidops
30	1872.2	60.0	34858	8	AB073165	Arabidops
31	1863.2	59.7	195837	8	ATCHRIV21	Arabidops
32	1841.8	59.0	85962	8	AB046431	Arabidops
33	1841.8	59.0	95463	8	F1809	Arabidops
34	1831.6	58.7	57000	8	AB046428	Arabidops
35	1794.4	57.5	111893	8	AC069555	Genomic S
36	1794.4	57.5	141430	8	AC069329	Genomic S
37	1793.2	57.4	123547	8	AC074109	Arabidops
38	1744.8	55.9	68639	8	AP001301	Arabidops
39	1739	55.7	105420	8	AC006268	Arabidops
40	1738	55.7	196517	8	ATCHRIV20	Arabidops
41	1738	55.7	28670	8	AC007311	Arabidops
42	1698.6	54.4	108227	8	ATF7P3	Arabidops
43	1665.2	53.3	115963	8	F28J9	Arabidops
44	1638.4	52.5	33104	6	AX059479	Sequence
45	1638.4	52.5	96988	8	ATF26B15	Arabidops

ALIGNMENTS

RESULT 1
LOCUS AX099282 3122 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 27 from Patent WO0120010.
ACCESSION AX099282
VERSION AX099282.1 GI:13538455
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1
REFERENCE Steimer, A., Mittelsten Scheid, O. and Paszkowski, J.
AUTHORS Transcriptionally silenced plant genes
TITLE

JOURNAL Patent: WO 0120010-A 27 22-MAR-2001;

Syngenta Participations AG (CH)

FEATURES Location/Qualifiers

source

1. .3122
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

ORIGIN

Query Match 100.0%; Score 3122; DB 6; Length 3122;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 TTCAATATTCGACCTCTCTCTCTCAATTCCTTGATCCAAAAGACACAAACGCGCCATC 60
Db 1 TTCAATATTCGACCTCTCTCTCTCAATTCCTTGATCCAAAAGACACAAACGCGCCATC 60
61 GCTTTCCTCACAACCTCTCACTGACCAACGCGCGCTCTCTCACTTACTCGGCTTCATCG 120
Db 61 GCTTTCCTCACAACCTCTCACTGACCAACGCGCGCTCTCTCACTTACTCGGCTTCATCG 120
121 CTCTCATCGCATCTCTCAACATCTCGACTCGCGATATCACTCGAGCTCGCGCTTCT 180
Db 121 CTCTCATCGCATCTCTCAACATCTCGACTCGCGATATCACTCGAGCTCGCGCTTCT 180
181 CACCGCTCTCCATCGTCAACGCTGCTCTCTCTCAAGGAAACAACTCGAGCTCTCC 240
Db 181 CACCGCTCTCCATCGTCAACGCTGCTCTCTCTCAAGGAAACAACTCGAGCTCTCC 240
241 ATTTCACTACTCGACTCTACCAACGCGGCTTCAACCACTCTAGCTTTAAACCACT 300
Db 241 ATTTCACTACTCGACTCTACCAACGCGGCTTCAACCACTCTAGCTTTAAACCACT 300
301 CGACCACTTCAACATCAACATCAATCGTTTCTCTCCATTAAGCTTTGACATCT 360
Db 301 CGACCACTTCAACATCAACATCAATCGTTTCTCTCCATTAAGCTTTGACATCT 360
361 CGACGCTGAACATCTTATCACTTCAAGCTCTCTCATCTCTTCTTCAACACCGCT 420
Db 361 CGACGCTGAACATCTTATCACTTCAAGCTCTCTCATCTCTTCTTCAACACCGCT 420
421 GCTCTCATCTCCCGAGAAAGCTTGTCACTCTCTCACTCATCACTCACTCGATTC 480
Db 421 GCTCTCATCTCCCGAGAAAGCTTGTCACTCTCTCACTCATCACTCACTCGATTC 480
481 AGCAACCAAACTCGACCTCGTCTCTTGGCACTCATAGTCACTCGATCTCTCTCACTCA 540
Db 481 AGCAACCAAACTCGACCTCGTCTCTTGGCACTCATAGTCACTCGATCTCTCTCACTCA 540
541 TCCTTCATCATCTCCCTTACTCGACCAACGCGTCTCGCTCCACCAATTCGCAATTTAAAG 600
Db 541 TCCTTCATCATCTCCCTTACTCGACCAACGCGTCTCGCTCCACCAATTCGCAATTTAAAG 600
601 CTCACTCGATTTGTAAGAGAGAGAGTGAAGCTCAACCAACGCGCTTCACTCGACGCTTT 660
Db 601 CTCACTCGATTTGTAAGAGAGAGAGTGAAGCTCAACCAACGCGCTTCACTCGACGCTTT 660
661 CCCTCTACACATTCACACTCGACCAACGCTGCTACCATCTCCACACCGCTCTTCTGTCAC 720
Db 661 CCCTCTACACATTCACACTCGACCAACGCTGCTACCATCTCCACACCGCTCTTCTGTCAC 720
721 CATACACTCGACCAACCTCTCAAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 CATACACTCGACCAACCTCTCAAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
781 CAGTTTCACTCAACCGGTTTACTCGACCGGTAGCTGTTTGAATGTTGTTTGGTTT 840
Db 781 CAGTTTCACTCAACCGGTTTACTCGACCGGTAGCTGTTTGAATGTTGTTTGGTTT 840
841 GCTATTACTCAACATTAAGTTTATCTTTGAGTTTCTGCTGTTTGAATGTTTCACTCATG 900
Db 841 GCTATTACTCAACATTAAGTTTATCTTTGAGTTTCTGCTGTTTGAATGTTTCACTCATG 900

901 AGTAACCTACAGTCGAAATCTCTATGGACCCCTGATTATTAATCTGGATGAAGCTAAGTCC 960
Db 901 AGTAACCTACAGTCGAAATCTCTATGGACCCCTGATTATTAATCTGGATGAAGCTAAGTCC 960
961 TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATATAGGATGAATTTGAA 1020
Db 961 TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATATAGGATGAATTTGAA 1020
1021 CGCTCTGCAGCTCGACGTAAATCAAGAGAGCTGAATTCCTAGAGAGAGAGAGGCGCATG 1080
Db 1021 CGCTCTGCAGCTCGACGTAAATCAAGAGAGCTGAATTCCTAGAGAGAGAGAGGCGCATG 1080
1081 TCGAGTAGATATGAGCTGATTGATGAGGATATCAAACTCAGATATAGCCAGAGCTCATGG 1140
Db 1081 TCGAGTAGATATGAGCTGATTGATGAGGATATCAAACTCAGATATAGCCAGAGCTCATGG 1140
1141 CGAAGGAGAGAGAGCTTACTGAAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGA 1200
Db 1141 CGAAGGAGAGAGAGCTTACTGAAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGA 1200
1201 TTCTTTGAGATGAATGACTTCTCGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCCGAG 1260
Db 1201 TTCTTTGAGATGAATGACTTCTCGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCCGAG 1260
1261 TTGGGGTTACTGGAGGAGCTGACATCTGTTGAGAGAGTGTCTATCTGATAAGGAGAGA 1320
Db 1261 TTGGGGTTACTGGAGGAGCTGACATCTGTTGAGAGAGTGTCTATCTGATAAGGAGAGA 1320
1321 CAATCGAGTTTCTTTCCACACTGCAAGTGGAAATGTATGAGGAGCTCACAGACTTTGAGC 1380
Db 1321 CAATCGAGTTTCTTTCCACACTGCAAGTGGAAATGTATGAGGAGCTCACAGACTTTGAGC 1380
1381 TGGATACCATGGGGTTAGGCTTTTGGAGTTCTTAGTGATGAACAGCGGTACCCAGATT 1440
Db 1381 TGGATACCATGGGGTTAGGCTTTTGGAGTTCTTAGTGATGAACAGCGGTACCCAGATT 1440
1441 AGATCAAGAAATTCGAAAGACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGGT 1500
Db 1441 AGATCAAGAAATTCGAAAGACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGGT 1500
1501 TTGACAGGAGAGCTTAAAGATTGTTGGGTACTATTTGGGAAACAACTACCGTAAACT 1560
Db 1501 TTGACAGGAGAGCTTAAAGATTGTTGGGTACTATTTGGGAAACAACTACCGTAAACT 1560
1561 CGACGCGGTCCAGAGAGCAACCAATCCGAGTCTGTGATTTCGCTTACTTACGCGCTCGG 1620
Db 1561 CGACGCGGTCCAGAGAGCAACCAATCCGAGTCTGTGATTTCGCTTACTTACGCGCTCGG 1620
1621 TTGCCAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTAAACACAGACATGAAGA 1680
Db 1621 TTGCCAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTAAACACAGACATGAAGA 1680
1681 TGATAGATTACAGCGTTATAGGGATTCTCCGCTTACAAAAGGAAAGAAATGTCTGTAGAG 1740
Db 1681 TGATAGATTACAGCGTTATAGGGATTCTCCGCTTACAAAAGGAAAGAAATGTCTGTAGAG 1740
1741 GAGATCTTAAACGATCAACCAACGATTAATCCCTGTTGATCCATCTGTGGTGGTACATGA 1800
Db 1741 GAGATCTTAAACGATCAACCAACGATTAATCCCTGTTGATCCATCTGTGGTGGTACATGA 1800
1801 AGTGGCGCTGACAAACGCGCAAGAGAGAGTAAAGAGGAGCACTATCGTGGGTGGGTTG 1860
Db 1801 AGTGGCGCTGACAAACGCGCAAGAGAGAGTAAAGAGGAGCACTATCGTGGGTGGGTTG 1860
1861 TGACGCCAATTCGAAAGTTTGTGGAGTTCCGCTCAAGAGAGTAGGGTTAGCACCGAGAA 1920
Db 1861 TGACGCCAATTCGAAAGTTTGTGGAGTTCCGCTCAAGAGAGTAGGGTTAGCACCGAGAA 1920
1921 TGATGGACTTGGATCACTTGGCCGATGTGAGTTCTCTGAGTTTGCATGTTGGGACT 1980
Db 1921 TGATGGACTTGGATCACTTGGCCGATGTGAGTTCTCTGAGTTTGCATGTTGGGACT 1980
1981 TTCACCGCTACAGGTTTCGAGCATTCATCGATTAGAAATCGCCAACTTCTTTTCCCTGCA 2040

Db	55567	GCCTTCCCTCACAACCTCTCACTCGACCACCGCCGCTCTCTCACTTACTCGGCTTCATCG	55508	Db	54487	CGCAAGGAGCAAGCTACTGAAACAAATCCGAGGTTTACAGTGGAGGAGTATATCAGA	54428
Qy	121	CTCTCATCGGCATCTCTCAACATACCTGACCTCGCGATATACCTCGAGCTCGCGCTTCT	180	Qy	1201	TTCTTTGAGATGAATGACTTCTCGGGAACGAGGTATCCCTGATATGATGACTTTAGCCGAG	1260
Db	55507	CTCTCATCGGCATCTCTCAACATACCTGACCTCGCGATATACCTCGAGCTCGCGCTTCT	55448	Db	54427	TTCTTTGAGATGAATGACTTCTCGGGAACGAGGTATCCCTGATATGATGACTTTAGCCGAG	54368
Qy	181	CACGGCTCTTCATCGTCAAGCTGCTGCTCTCTCCAAAGGAACAACTCGAGCTCTCC	240	Qy	1261	TTGGGGTTACTCGGAGGACGTGACAGCATCTCTTCGAGAAGTGTCTATCTGATAAGGAGAGA	1320
Db	55447	CACGGCTCTTCATCGTCAAGCTGCTGCTCTCTCCAAAGGAACAACTCGAGCTCTCC	55388	Db	54367	TTGGGGTTACTCGGAGGACGTGACAGCATCTCTTCGAGAAGTGTCTATCTGATAAGGAGAGA	54308
Qy	241	ATTTCATCTACCTGACCTCTACCAACGAGCGCTTCAACACTTCTAGCTCTTAAACACT	300	Qy	1321	CAATCGAGTTTCTTCCACTCCAACTGTAAGTGAATGTATGAGGACTCACAGACTTTTGAGC	1380
Db	55387	ATTTCATCTACCTGACCTCTACCAACGAGCGCTTCAACACTTCTAGCTCTTAAACACT	55328	Db	54307	CAATCGAGTTTCTTCCACTCCAACTGTAAGTGAATGTATGAGGACTCACAGACTTTTGAGC	54248
Qy	301	CGACCACTTTCACCATCAACCAATCAAAATCGTTTCTCTCCATTAAGCTTTCACATCT	360	Qy	1381	TGATACCACTGGGGTTAGGCTTCTGACGTTCTTAGTGTGATGAACAGCGGTACACAGTTT	1440
Db	55327	CGACCACTTTCACCATCAACCAATCAAAATCGTTTCTCTCCATTAAGCTTTCACATCT	55268	Db	54247	TGATACCACTGGGGTTAGGCTTCTGACGTTCTTAGTGTGATGAACAGCGGTACACAGTTT	54188
Qy	361	CGACCGCTGAACACTTATACCTTCAAGCTCTCATCTCTTCATCGTTTCCAAACCGCT	420	Qy	1441	AGATCAAGAAATTTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGGT	1500
Db	55267	CGACCGCTGAACACTTATACCTTCAAGCTCTCATCTCTTCATCGTTTCCAAACCGCT	55208	Db	54187	AGATCAAGAAATTTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGGT	54128
Qy	421	GCTCTCATCCCCACGAAAGCTTGTATCACTCTCACTCTCACTCATCAGAGTTTCACTCGATTC	480	Qy	1501	TTGACAGGGAAGAGCTTAAAGGATTTGTGGCTACTATTGGGAACAATCTACCGCTTAAACT	1560
Db	55207	GCTCTCATCCCCACGAAAGCTTGTATCACTCTCACTCTCACTCATCAGAGTTTCACTCGATTC	55148	Db	54127	TTGACAGGGAAGAGCTTAAAGGATTTGTGGCTACTATTGGGAACAATCTACCGCTTAAACT	54068
Qy	481	AGTAACCAAACTCGACTCGTCTCTCTTGGCACTCATAGTCACTCGATCTCTCTCCACCA	540	Qy	1561	CGACCGGTCCTCAAGAGCAACCAATCCGAGTCTCTGTGATTCGCTACTTTAGCGCTCGG	1620
Db	55147	AGTAACCAAACTCGACTCGTCTCTCTTGGCACTCATAGTCACTCGATCTCTCTCCACCA	55088	Db	54067	CGACCGGTCCTCAAGAGCAACCAATCCGAGTCTCTGTGATTCGCTACTTTAGCGCTCGG	54008
Qy	541	TCTTCATCATCTCCCTTACTCGACACCGTGTGCTCTCGCTCCACCAATGCAATTTAAAG	600	Qy	1621	TTGCCAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTTAACAGACATCAAGA	1680
Db	55087	TCTTCATCATCTCCCTTACTCGACACCGTGTGCTCTCGCTCCACCAATGCAATTTAAAG	55028	Db	54007	TTGCCAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTTAACAGACATCAAGA	53948
Qy	601	CTCACTCGATTTGTCAAAGAGAGAGAGTGAAGTCAACCGCGCTCTGCTCCACCAATGCAATTTAAAG	660	Qy	1681	TGATAGTTTACGCTTATAGGATTTCTCCGCTTACAAAGGAAAGAAATGTCTCTGAGAG	1740
Db	55027	CTCACTCGATTTGTCAAAGAGAGAGAGTGAAGTCAACCGCGCTCTGCTCCACCAATGCAATTTAAAG	54968	Db	53947	TGATAGTTTACGCTTATAGGATTTCTCCGCTTACAAAGGAAAGAAATGTCTCTGAGAG	53888
Qy	661	CCCTCTACACATTTCACTCGACCAACCGTGTGCTTACCATCTCCACACCCGCTCTTGTTCAC	720	Qy	1741	GAGATCTTAAACACTCACCAACAGTAATGCCCTCTGTGTGATTCATCTGTGTGGGTACATGA	1800
Db	54967	CCCTCTACACATTTCACTCGACCAACCGTGTGCTTACCATCTCCACACCCGCTCTTGTTCAC	54908	Db	53887	GAGATCTTAAACACTCACCAACAGTAATGCCCTCTGTGTGATTCATCTGTGTGGGTACATGA	53828
Qy	721	CATACACTCGACCAACACTCTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG	780	Qy	1801	AGTGGCGCTGACAAACGCAAGAGAGTAAAGAGGACACTATCGTGGGTGGGCTTG	1860
Db	54907	CATACACTCGACCAACACTCTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG	54848	Db	53827	AGTGGCGCTGACAAACGCAAGAGAGTAAAGAGGACACTATCGTGGGTGGGCTTG	53768
Qy	781	CAGTTTCACTCAACCGGTTTACTCGACCGGTAGCTGGTTTGTAGATTTGTTTGGTTTT	840	Qy	1861	TGACGCCAATTTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGAGAA	1920
Db	54847	CAGTTTCACTCAACCGGTTTACTCGACCGGTAGCTGGTTTGTAGATTTGTTTGGTTTT	54788	Db	53767	TGACGCCAATTTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGAGAA	53708
Qy	841	GCTATTACTAATTAACGTTTATCTTTGAGTTTCTGCTCTGTTTGTAGTTTTCATCATG	900	Qy	1921	TGATCGACTTGGATCACCTTGGCCGATGTGATTTCTCTGAGTTTGCATGTTGGGACT	1980
Db	54787	GCTATTACTAATTAACGTTTATCTTTGAGTTTCTGCTCTGTTTGTAGTTTTCATCATG	54728	Db	53707	TGATCGACTTGGATCACCTTGGCCGATGTGATTTCTCTGAGTTTGCATGTTGGGACT	53648
Qy	901	AGTAAGTACAGTGAAGTAACTCTATGACCTGCTGATTAATGATGATGATGATGATGATG	960	Qy	1981	TTACCGCTTACAGGTTTCGAGCATTCATCGATTTAGAAATCGCCAAACATCTTTTCCCTGCA	2040
Db	54727	AGTAAGTACAGTGAAGTAACTCTCTATGACCTGCTGATTAATGATGATGATGATGATG	54668	Db	53647	TTACCGCTTACAGGTTTCGAGCATTCATCGATTTAGAAATCGCCAAACATCTTTTCCCTGCA	53588
Qy	961	TGGTCCACTAGACCGGAGTGAAGCAACATGTTTACGAGAGCTATAGGATGAATTTGAA	1020	Qy	2041	TTTACGCTTACAGGTTTCGAGGCTAGGAAATTTGACTTCAAGCTCTCGCTTGAAGATC	2100
Db	54667	TGGTCCACTAGACCGGAGTGAAGCAACATGTTTACGAGAGCTATAGGATGAATTTGAA	54608	Db	53587	TTTACGCTTACAGGTTTCGAGGCTAGGAAATTTGACTTCAAGCTCTCGCTTGAAGATC	53528
Qy	1021	CGCTCTGAGCTCGAGCTGAATCAAGAGAGCTGAAATCGCTTAGAGGAAGAGGCGGATG	1080	Qy	2101	TTTATTTTCGAGGCTAGGCTAGGAAATTTGACTTCAAGCTCTCGCTTGAAGATC	2160
Db	54607	CGCTCTGAGCTCGAGCTGAATCAAGAGAGCTGAAATCGCTTAGAGGAAGAGGCGGATG	54548	Db	53527	TTTATTTTCGAGGCTAGGCTAGGAAATTTGACTTCAAGCTCTCGCTTGAAGATC	53468
Qy	1081	TCGAGTATGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1140	Qy	2161	AAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2220
Db	54547	TCGAGTATGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	54488	Db	53467	AAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	53408
Qy	1141	CGCAAGGAGAGAGCTACTGACAAATCCGACGAGTTTACAGTGGAGGAGTATATCAGA	1200	Qy	2221	TCAGTGGAGCATATACCTCCAGCGAGGAAAGCAAGAGTTTTCAGCGAAGCTCACAGGAACA	2280
Db				Db	53407	TCAGTGGAGCATATACCTCCAGCGAGGAAAGCAAGAGTTTTCAGCGAAGCTCACAGGAACA	53348

QY	2281	ACAGCAAGCTGCAGAGTGGTGCAGAAACAGAGATAGTTACTCGCAAGTGCCTCAGGG	2340
DB	53347	ACAGCAAGCTGCAGAGTGGTGCAGAAACAGAGATAGTTACTCGCAAGTGCCTCAGGG	53288
QY	2341	CTATCAAGTTCTGAGGACAGATCAGCTGCTCTCTCCACTACAACTATTCCGCAAT	2400
DB	53287	CTATCAAGTTCTGAGGACAGATCAGCTGCTCTCTCCACTACAACTATTCCGCAAT	53228
QY	2401	GACAGTCCCTCAGGACATGCTTTCGAGGAGATATGACGGCCCGAGCTTAGAGAGAGA	2460
DB	53227	GACAGTCCCTCAGGACATGCTTTCGAGGAGATATGACGGCCCGAGCTTAGAGAGAGA	53168
QY	2461	AGATTCTGCATGCTCCCTCGAGGAGCATTCATCTTCGAGGCTCGTGAATCTAGGAAGAATA	2520
DB	53167	AGATTCTGCATGCTCCCTCGAGGAGCATTCATCTTCGAGGCTCGTGAATCTAGGAAGAATA	53108
QY	2521	GGAGAACGACACTCACTCGATCTAGCAGGAGGACAGACGACTTCTCGAGTCTCGTAGTT	2580
DB	53107	GGAGAACGACACTCACTCGATCTAGCAGGAGGACAGACGACTTCTCGAGTCTCGTAGTT	53048
QY	2581	TACCGACCGGCTGCTGGCGGCAATAGAGAGAGGTCGAGTATCTTCAGAGCGGTG	2640
DB	53047	TACCGACCGGCTGCTGGCGGCAATAGAGAGAGGTCGAGTATCTTCAGAGCGGTG	52988
QY	2641	CTGCGCCGCCACAGAGCTGATGAGATCGAGTACCCACATGCTGGAGCTGATACGGAACATG	2700
DB	52987	CTGCGCCGCCACAGAGCTGATGAGATCGAGTACCCACATGCTGGAGCTGATACGGAACATG	52928
QY	2701	GCGGTTCTGCTATGCTGGGAGCAATCAGAGGAGGCAATGATACCACTTCGTTCTCAT	2760
DB	52927	GCGGTTCTGCTATGCTGGGAGCAATCAGAGGAGGCAATGATACCACTTCGTTCTCAT	52868
QY	2761	TATTCGACTGAGTAAAGCGCTCACTTCAACATATATATATATATATATATATATATATAT	2820
DB	52867	TATTCGACTGAGTAAAGCGCTCACTTCAACATATATATATATATATATATATATATATAT	52808
QY	2821	TCTTTATTTCTTTTCAGTATGATTTGCTGAGTACTCTCTTCCAAAGTTTATTCACA	2880
DB	52807	TCTTTATTTCTTTTCAGTATGATTTGCTGAGTACTCTCTTCCAAAGTTTATTCACA	52748
QY	2881	CAGTGGACTGTGATTTAAGTTTGGGGAGGCTCAGAGAGTATGTTGCAATGTATATA	2940
DB	52747	CAGTGGACTGTGATTTAAGTTTGGGGAGGCTCAGAGAGTATGTTGCAATGTATATA	52688
QY	2941	TTTTTAAGTCTGCATTCATTAAGCATAGAAAACCAAAAATAAATAAATTTTACA	3000
DB	52687	TTTTTAAGTCTGCATTCATTAAGCATAGAAAACCAAAAATAAATAAATTTTACA	52628
QY	3001	AAATGATTTCAAAAAAGAGTGTTCATGTAGTTGCAATTCATTTAGGATCAAGTCTAG	3060
DB	52627	AAATGATTTCAAAAAAGAGTGTTCATGTAGTTGCAATTCATTTAGGATCAAGTCTAG	52568
QY	3061	AGTGTTCATTTAGGATTTGTCATATGATAGGAGGATATGATAGATAGCTTTGTAAG	3120
DB	52567	AGTGTTCATTTAGGATTTGTCATATGATAGGAGGATATGATAGATAGCTTTGTAAG	52508
QY	3121	CA 3122	
DB	52507	CA 52506	

RESULT 3
 AF058825
 LOCUS AF058825 110157 bp DNA linear PLN 15-APR-1998
 DEFINITION Arabidopsis thaliana BAC F7N22.
 ACCESSION AF058825
 VERSION AF058825.1 GI:3047060
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 110157)
 Washington University Genome Sequencing Center.
 The A. thaliana Genome Sequencing Project
 Unpublished (1997)
 2 (bases 1 to 110157)
 Dante, M.
 The sequence of A. thaliana F7N22
 Unpublished (1998)
 3 (bases 1 to 110157)
 Waterston, R.
 Direct Submission
 Submitted (09-APR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63108, USA
 e-mail: r.watson@wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
 M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
 by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The actual start of this clone is at base position 1 of F7N22;
 actual end is at 110157 of F7N22.

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES
 source
 1..110157
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="IV"
 /map="unknown"
 /clone="F7N22"
 3032..3821
 /note="Pseudogene; similar to reverse transcriptases"
 complement(6554..8653)
 /gene="F7N22.18"
 complement(join(6554..6961,7363..7584,8127..8253,
 8487..8653))
 /gene="F7N22.18"
 /codon_start=1
 /evidence=not experimental
 /protein_id="AAC13576.1"
 /db_xref="GI:3047062"
 /translation="MGLLLSKINPLKASDNHNPISRFNNLTLDLFDLSPGHASLT
 RLDPNRPGLTFNEDSSVVGVCIFERDFDPHAYWTTPQVITCWETPACRKT
 EETTPDLVLDTHKADRTFDGKSQIYKEVSRABEETQLCGSSSESIASGG
 LSVQAMNITYTHVAPKKRIYVVGSLHLEASHTCLPLPRDDPVLLEKLAASAC
 IANQAEKINSFGVYFDYADKDPVFAAI FRGASSTRTKPTSANPRPEVANL

misc_feature

gene

CDS

misc_feature
 11245..114802
 /note="Pseudogene; similar to reverse transcriptases"
 gene
 27112..28030
 /gene="F7N22_3"
 CDS
 join(27112..27748,27822..28030)
 /gene="F7N22.3"
 /note="contains similarity to peptidyl-prolyl cis-trans isomerase [Pfam: pro isomerase.hmm, score: 23.86 and 28.4]; coded for by A. thaliana cDNA N97251; coded for by A. thaliana cDNA T43563"
 /codon_start=1
 /protein_id="AACI3578.1"
 /db_xref="GI:3047064"
 /translation="MASSIIIPILITRRNLISLTIAVTTPPPAKPPSPDITIDR VFLDFSLCTPFRSDPSATPLSCSDSTPLGRVVLGLYGRHVPIIVTSFKRMCTSS STSYKNTPVHKIFPGQYLACGQGSGRDTAEGVSUARDLRNDTVDVNSKAFLPHAR AGVVLSICLSENNDDDIRLDPYRNVEFLITPGPSPQLDGGNIVFGTGLEGLDVYT SISSPTPKSENIKQFNDAFEUGDERAQNARSLLNRPLKTIVFISGCGLKVTPSL SPTLP"
 32752..33612
 /gene="F7N22.4"
 CDS
 join(32752..32937,33015..33103,33358..33424,33502..33612)
 /gene="F7N22.4"
 /note="contains similarity to human OS-9 precursor (GB:U41635)"
 /codon_start=1
 /evidence=not experimental
 /protein_id="AACI3579.1"
 /db_xref="GI:3047065"
 /translation="MVMLDRHGNFLCYLPKEKATSGWTSQQNISVMMETQQLVK LKTDPELQISEKLRFQEGMWSEYFCQXKYQLHVEDENKGVFYWHSHVYTNGT TCMTGPSREYVEFVCAEATRAMVTSITELSTCKYALTVCQPTLCCKP"
 complement(35192..39289)
 /note="Pseudogene; similar to reverse transcriptases"
 complement(39286..40766)
 /gene="F7N22.17"
 CDS
 complement(join(39286..40383,40614..40766))
 /gene="F7N22.17"
 /note="similar to Arabidopsis thaliana retrotransposon Ta1-1 (GB:I47193)"
 /codon_start=1
 /evidence=not experimental
 /protein_id="AACI3575.1"
 /db_xref="GI:3047061"
 /translation="MAMLSGDFTGDFVPVLHOLISPRWSAFETIGIRFFLFNFDOGY SLPSNYAAVYNRSEYVTVTPRKQNLRALIQMRVWGFPDSCVRIDERGRVOFK PQTEAMNVLVTRGPMSFDNMULSIHRWYNISTELKIIPFWQITGIPLLFNAM ARCNGRLGVTVDFDENSHSVFRVKIDNLDLDFRQFVQFAVDENTVIKYRF ERLNFSKCGSLKHDKCVLAFFDDBPAEISDDNDODNOGDNKDMDSDTLQ TVDPATIPGITPTOPFGNKGPNTQTTFVSPSTDITLVERLYHLAKLARVT QPTSLNDLLEDSDNAQOQVLLKRKRVQEDMYRKEADENAVLSQLCKKERKABS AGCSIIHGIDGGAGVPPRPDI"
 46610..47884
 /note="Pseudogene; similar to retrovirus-related POL polyproteins"
 57941..59852
 /gene="F7N22_6"
 CDS
 join(57941..58149,58209..59137,59251..59852)
 /gene="F7N22.6"
 /note="contains similarity to retrovirus-related POL polyproteins"
 /codon_start=1
 /evidence=not experimental
 /protein_id="AACI3580.1"
 /db_xref="GI:3047066"
 /translation="MTDINTNSKIVGNTTWETPDVALHLAKELEKTOIGSGMDNV RRLFANDSTNTTIPSSASOMMPDIFNNERYLTDEPTVPQGTIVTGGMGTWAQ GDYCKGLILNRLVLDLYNKAASKTLLTLENKYKTDESQQRSTAFELNFKM VDSKPMEQVEALQRISETLELGMGINCNVFKINCLIEKLPPGLWDLFKNLNFKKAM TEDDLVRLMIEGNYGAHAQAQGHVDVAHKAUKLGKGKFSIPQNLKVSST NFKKSNERPFKCCHICKGHDVCKSKSQAANTTEEDMAVATEECNVMDV FWFYNYDTCTGTHICTDTMFSTYYVNKSNEOLFMGNTAMSIEGGLVMGSMTVEP

KNVAGASVVEMNNENITVLVESFTTWHRLGHVNYKIMRQNNMLIKPKFTNQEKCE
VCHAKLTKTPSPRVERTEPNMLIKFKTNQEKCEL CVQAKLTKTPSPRVERTEPL
GLIYDLCDLCKYVQVRGGKKFPVTFIDDCTRYVYVLLSHKDEALVKFKFTLEVNG
LQRTIKIVRSDRG"
63170..64252
/gene="F7N22.7"
63170..64252
/gene="F7N22.7"
/note="similar to Arabidopsis thaliana retrotransposon
Achila (GB:X81801)"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13581.1"
/db_xref="GI:3047067"
/translation="MEALNTKRVYLVLEGOSASTSAPKVTGLPGKSIONKPEYATAHAIT
ICHDELPTRHVPDLIIIGNDVQGEASTQLEVSVEFNHSDGPHLIQTSSEKAAT
IKRIKVRKPTPLPSRALPWTFRKAWERYKSAATKQLDEAVNPLMEVLYLLHDPH
KDVNNLIERINMYQDLDDESATRSAAKRIIVOKNLEDGSLTFLGSGELAFSDC
LDCIAGVSMPLSVARRLEIYQKPCDPLTILADRSYRKFPGMLKDLPMVINGVEVP
TENVFLDMEVSHKDFLILGRPFPLASGVGDVIREGKISLNLGKHKLQFDINKIPQES
TEDEKTSKDYRVIPEGVEYETEKIKKA"
64919..67482
/gene="F7N22.8"
Join(64919..64986,66088..66270,66336..66473,66723..67482)
/gene="F7N22.8"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13577.1"
/db_xref="GI:3047063"
/translation="MIQNKETDEGPTAAASPTLPTRDEFERSAARNQRRRAFIARGK
RAMSRVELIDEDIKTEYESPESMRKETKLKNSDEVTPGVGTGGRAASVREVS
KEETIEFLTLQVMYEDFLDITMGESTGTVDNMDKMDISALIGLRLTKGNV
LRGLNDSPVPMPLIHLGCVGMKWLNGKKVGRGALCVGVVTPILKVCVGPLKEVG
LAPMLDHLRLRCERSEFDMVGDFHYREPHSSIRIANILLFPCIYATRILGSRNIDF
KPALEDLYFEGSPTEISHTEGATDVEDTIDIDBAEDTSMYHFSHEIHPARKSK
SLSAHRNNSKLQAKKKQDLAKCLRALFKLKDICSSTTTTPQ"
72089..73326
/note="Pseudogene; similar to retrovirus-related POL
polyproteins"
complement(84219..86759)
/gene="F7N22.15"
complement(join(84219..84649,84749..85148,85424..85967,
86365..86759))
/gene="F7N22.15"
/note="contains similarity to retrotransposon-like

Query Match	100.0%;	Score 3122;	DB 8;	Length 110157;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 3122;	Conservative 0;
Qy 1	TTCATATATTCGACCTCTTCTTCATCTTCGTGATCCAAAGACACAAACGCGGCATC	60		
Db 65081	TTCATATATTCGACCTCTTCTTCATCTTCGTGATCCAAAGACACAAACGCGGCATC	65140		
Qy 61	GCTTTCCTCCACAACTCTCACTCGACACGCGCGCTCTCTCACTTACTCGGCTTCATCG	120		
Db 65141	GCITTCCTCCACAACTCTCACTCGACACGCGCGCTCTCTCACTTACTCGGCTTCATCG	65200		
Qy 121	CTCTCATCGGCATCTCTCAACATACCTCGACCTCGGATATCACTCGAGTCGCGCTTCT	180		
Db 65201	CTCTCATCGGCATCTCTCAACATACCTCGACCTCGGATATCACTCGAGTCGCGCTTCT	65260		
Qy 181	CACGGCTCTCCATCGTCACGGCTGCTCCCTCTCTCCAGGAACAACTCGAGCTCTCC	240		
Db 65261	CACGGCTCTCCATCGTCACGGCTGCTCCCTCTCTCCAGGAACAACTCGAGCTCTCC	65320		
Qy 241	ATTTCATCACTCGACCTCTACACCAACGCGCTTCACCACTTCTAGCTCTTAACCACT	300		
Db 65321	ATTTCATCACTCGACCTCTACACCAACGCGCTTCACCACTTCTAGCTCTTAACCACT	65380		
Qy 301	CGACCACTTCACCATCAACCAATCAAAATCGTTTCTCTCATTAAGCTTGCATCACT	360		
Db 65381	CGACCACTTCACCATCAACCAATCAAAATCGTTTCTCTCATTAAGCTTGCATCACT	65440		

Qy	361	CGACCGCTGAACACTTATCACCCTTCAAGCTCCCTCATCTCTTATCGTTTCCAAACACCGCT	420	1441	AGATCAAGAAATTGGAGAACTGTTTGGTTTCCCTAGTGGAAAGGGAACCAACCCAGGT	1500
Db	65441	CGACCGCTGAACACTTATCACCCTTCAAGCTCCCTCATCTCTTATCGTTTCCAAACACCGCT	65500	65521	AGATCAAGAAATTGGAGAACTGTTTGGTTTCCCTAGTGGAAAGGGAACCAACCCAGGT	66580
Qy	421	GCTCTCATCCCGACGAAAGCTTGTCTCATCTCTCATCTCATCAACAGTTTCACTCGATTTC	480	1501	TTGACAGGGAGAGCTTAAAGATTGTGGGCTACTATTGGGAAACAATCTACCGTTAACT	1560
Db	65501	GCTCTCATCCCGACGAAAGCTTGTCTCATCTCTCATCTCATCAACAGTTTCACTCGATTTC	65560	65581	TTGACAGGGAGAGCTTAAAGATTGTGGGCTACTATTGGGAAACAATCTACCGTTAACT	66640
Qy	481	AGCAACCAAACTCGAAGCTGCTCTCTTTGGCACTCATAGTCACTCGATCTCTCTCACCA	540	1561	CGACCGGCTCCAGAGCAACCAAAATCCGGAGTCTCTGATTTGCTTACGCTTTCAGCGCTCGG	1620
Db	65561	AGCAACCAAACTCGAAGCTGCTCTCTTTGGCACTCATAGTCACTCGATCTCTCTCACCA	65620	66641	CGACCGGCTCCAGAGCAACCAAAATCCGGAGTCTCTGATTTGCTTACGCTTTCAGCGCTCGG	66700
Qy	541	TCCTTCATCATCTCTTACTCGACACCGTGGTCTCGCTCCACCATTCGCAATTTAAAG	600	1621	TTGCCAATTTTTTTTACTCCAGGAGTCTACAGGACCGTCTCTAAACACAGACATGAAGA	1680
Db	65621	TCCTTCATCATCTCTTACTCGACACCGTGGTCTCGCTCCACCATTCGCAATTTAAAG	65680	66701	TTGCCAATTTTTTTTACTCCAGGAGTCTACAGGACCGTCTCTAAACACAGACATGAAGA	66760
Qy	601	CTCACTCGATTGTCAAAGAGAAAGAGTGAAGTCAACACCGCACTCGACCGCGTTT	660	1681	TGATAGATTACGAGCTTATAGGGATTCTCCGCTTTACAAAAGGAAAGAAATGCTCTGAGAG	1740
Db	65681	CTCACTCGATTGTCAAAGAGAAAGAGTGAAGTCAACACCGCACTCGACCGCGTTT	65740	66761	TGATAGATTACGAGCTTATAGGGATTCTCCGCTTTACAAAAGGAAAGAAATGCTCTGAGAG	66820
Qy	661	CCCTCTACATTAACACTCGACACCGTGTCTACCATCTCCACACCGCTCTTGTTCAC	720	1741	GAGATCTTAAACGACTCACCAACGCAAGAGAGTAAGAGGAGCACTATGCGTGGGTGGCGTTG	1800
Db	65741	CCCTCTACATTAACACTCGACACCGTGTCTACCATCTCCACACCGCTCTTGTTCAC	65800	66821	GAGATCTTAAACGACTCACCAACGCAAGAGAGTAAGAGGAGCACTATGCGTGGGTGGCGTTG	66880
Qy	721	CATACACTCGACCAACACTCTCAAGTAAAGAAAAAAGAAAAAAGTCAAAACCGA	780	1801	AGTGGGCGCTGACAAAACGCAAGAGAGTAAGAGGAGCACTATGCGTGGGTGGCGTTG	1860
Db	65801	CATACACTCGACCAACACTCTCAAGTAAAGAAAAAAGAAAAAAGTCAAAACCGA	65860	66881	AGTGGGCGCTGACAAAACGCAAGAGAGTAAGAGGAGCACTATGCGTGGGTGGCGTTG	66940
Qy	781	CAGTTTCACTCAACCGGTTTACTCGACCGGTACGCTGGTTTATAGTTGTGTTTTTGGTTTT	840	1861	TGACGCCAATCTGAAAAGTTTGTGGAGTTCGCTCAAGAAAGTAGGGTTAGCACCGAGAA	1920
Db	65861	CAGTTTCACTCAACCGGTTTACTCGACCGGTACGCTGGTTTATAGTTGTGTTTTTGGTTTT	65920	66941	TGACGCCAATCTGAAAAGTTTGTGGAGTTCGCTCAAGAAAGTAGGGTTAGCACCGAGAA	67000
Qy	841	GCTATTACTAATTAACGTTTATCTTGAAGTTTCTGCTGTTTATAGTTTTCATCATG	900	1921	TGATCGACTTGGATCACTTGGCGCGATGTGAGTTCTCTGAGTTTGACATGGTGGCGACT	1980
Db	65921	GCTATTACTAATTAACGTTTATCTTGAAGTTTCTGCTGTTTATAGTTTTCATCATG	65980	67001	TGATCGACTTGGATCACTTGGCGCGATGTGAGTTCTCTGAGTTTGACATGGTGGCGACT	67060
Qy	901	AGTAACACTAGTGGAAATCTCTATGGACCTGATTATAATGTGGATGAAGTCAAGTCC	960	1981	TTTATTTTCAGGGGAGTCCGCCAACTGAGGAGATTAGTTCACACCGAGAGCTACAAATAG	2040
Db	65981	AGTAACACTAGTGGAAATCTCTATGGACCTGATTATAATGTGGATGAAGTCAAGTCC	66040	67061	TTTATTTTCAGGGGAGTCCGCCAACTGAGGAGATTAGTTCACACCGAGAGCTACAAATAG	67120
Qy	961	TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGATGAATTTGAA	1020	2041	TTTATTTTCAGGGGAGTCCGCCAACTGAGGAGATTAGTTCACACCGAGAGCTACAAATAG	2100
Db	66041	TGGTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGATGAATTTGAA	66100	67121	TTTATTTTCAGGGGAGTCCGCCAACTGAGGAGATTAGTTCACACCGAGAGCTACAAATAG	67180
Qy	1021	CGCTCTGAGCTGAGCTTAATCAAGAGAGCTGAATCGCTAGAGGAAAGAGGGGATG	1080	2101	TTTATTTTCAGGGGAGTCCGCCAACTGAGGAGATTAGTTCACACCGAGAGCTACAAATAG	2160
Db	66101	CGCTCTGAGCTGAGCTTAATCAAGAGAGCTGAATCGCTAGAGGAAAGAGGGGATG	66160	67181	TTTATTTTCAGGGGAGTCCGCCAACTGAGGAGATTAGTTCACACCGAGAGCTACAAATAG	67240
Qy	1081	TCGAGTAGATATGAGCTGATTGATGAGGATATCAAACTGAGTATGAGCCAGAGTCAATG	1140	2161	AAAGTTTTCATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCAT	2220
Db	66161	TCGAGTAGATATGAGCTGATTGATGAGGATATCAAACTGAGTATGAGCCAGAGTCAATG	66220	67241	AAAGTTTTCATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCAT	67300
Qy	1141	CGCAAGGAGACGAAGCTTACTGAACAAATCCGACGAGTTTACGAGGAGTATATCAGA	1200	2221	TCAGTGAGCATATACCTCCAGCGAGGAAAGCAAGAGTTTGGCGGAGCTCACAGGAACA	2280
Db	66221	CGCAAGGAGACGAAGCTTACTGAACAAATCCGACGAGTTTACGAGGAGTATATCAGA	66280	67301	TCAGTGAGCATATACCTCCAGCGAGGAAAGCAAGAGTTTGGCGGAGCTCACAGGAACA	67360
Qy	1201	TTCTTTGAGATGATGACTTCTGGGGAACGAGCTATCCCTGATATGAGACTTTAGCCAG	1260	2281	ACAGGAAGCTGCAAGAGTGTGCAAGAAACAGAGTAAGTTACTCGCCAGTGTCTCAGGG	2340
Db	66281	TTCTTTGAGATGATGACTTCTGGGGAACGAGCTATCCCTGATATGAGACTTTAGCCAG	66340	67361	ACAGGAAGCTGCAAGAGTGTGCAAGAAACAGAGTAAGTTACTCGCCAGTGTCTCAGGG	67420
Qy	1261	TTGGGGTTTACTGGAGGACGTGACGATCTCTTTCGAGAAAGTGTCTATCTGATAAGGAGGA	1320	2341	CTATCAAGTTTCTGAAAGGACCAAGATCAGCTGCTCTCTTCCACTACAACTATTTCCGCAAT	2400
Db	66341	TTGGGGTTTACTGGAGGACGTGACGATCTCTTTCGAGAAAGTGTCTATCTGATAAGGAGGA	66400	67421	CTATCAAGTTTCTGAAAGGACCAAGATCAGCTGCTCTCTTCCACTACAACTATTTCCGCAAT	67480
Qy	1321	CAATCGAGTTTCTTTCACACTGCAAGTGAATGATAGGAGACTCACAGACTTTGAGC	1380	2401	GACAGCTCCCTCAGGACATGCTTTCGAGGAGATATGACGCGCCGAGCTTAGAGAGAGA	2460
Db	66401	CAATCGAGTTTCTTTCACACTGCAAGTGAATGATAGGAGACTCACAGACTTTGAGC	66460	67481	GACAGCTCCCTCAGGACATGCTTTCGAGGAGATATGACGCGCCGAGCTTAGAGAGAGA	67540
Qy	1381	TGGATACCATGGGGTTAGGCTTCTTGTAGTTCTTGTAGTGAACACGCGGTACAGATTT	1440	2461	AGATTTTGTGATGTCCTCGGAGGCTTCTATTTTCGAGCTCTGTAATCTTAGGAAGAATA	2520
Db	66461	TGGATACCATGGGGTTAGGCTTCTTGTAGTTCTTGTAGTGAACACGCGGTACAGATTT	66520	67541	AGATTTTGTGATGTCCTCGGAGGCTTCTATTTTCGAGCTCTGTAATCTTAGGAAGAATA	67600
				2521	GGAGAACGACACTCACTCGATCTAGCAGCAGGAGCAGACGACTTCTCGAGTCTCGTAGTT	2580


```

Db      67601  GGAGAACGACACTCACTGATCTAGCAGCAGAGCAGAGCTTCTGCAGTCTCGTAGTT 67660
Qy      2581  TACGCGACCGCGGTGCTGCGCGCAATAGAAAGAGAGGTGAGTATCTCTCAGAGCGGTG 2640
Db      67661  TACGCGACCGCGGTGCTGCGCGCAATAGAAAGAGAGGTGAGTATCTCTCAGAGCGGTG 67720
Qy      2641  CTGCGCGCCACAGAGCTGATGAGATCGAGTACCGACATGCTGGAGCTGATACGGAACATG 2700
Db      67721  CTGCGCGCCACAGAGCTGATGAGATCGAGTACCGACATGCTGGAGCTGATACGGAACATG 67780
Qy      2701  GCGGTTCGCTATGCTGCTGGAGCAATCACAGGCGACCATGACTACCACTTCGTTTCAT 2760
Db      67781  GCGGTTCGCTATGCTGCTGGAGCAATCACAGGCGACCATGACTACCACTTCGTTTCAT 67840
Qy      2761  TATTCGACTGAGGTAAGCGCTCACTTCACCATATATATATATATATATATATATATATAT 2820
Db      67841  TATTCGACTGAGGTAAGCGCTCACTTCACCATATATATATATATATATATATATATATAT 67900
Qy      2821  TCTTATTTTGTTCAGTGATGATGATGCTCCTGAGTACTCTCTTCCAAAGTTTATTCACA 2880
Db      67901  TCTTATTTTGTTCAGTGATGATGATGCTCCTGAGTACTCTCTTCCAAAGTTTATTCACA 67960
Qy      2881  CAGTGGACTGTGATTTAAGTTTGGGGAGGGCTCAGGAGTATGTTGCAATTCATATATA 2940
Db      67961  CAGTGGACTGTGATTTAAGTTTGGGGAGGGCTCAGGAGTATGTTGCAATTCATATATA 68020
Qy      2941  TTTTAACTGCTGATTCATCTAAGGCATAGAAAACCAAAAAAAATTTAAAAATTTTCCAG 3000
Db      68021  TTTTAACTGCTGATTCATCTAAGGCATAGAAAACCAAAAAAAATTTAAAAATTTTCCAG 68080
Qy      3001  AAATGATTTACAAAAAAGAGTGTTCATGATGCTGATGCTGATGCTGATGCTGATGCTG 3060
Db      68081  AAATGATTTACAAAAAAGAGTGTTCATGATGCTGATGCTGATGCTGATGCTGATGCTG 68140
Qy      3061  AGTGTTCATTTAGGATTTGTCATATGATGCTAGGGGATATGATGATGATGATGATGAT 3120
Db      68141  AGTGTTCATTTAGGATTTGTCATATGATGCTAGGGGATATGATGATGATGATGATGAT 68200
Qy      3121  CA 3122
Db      68201  CA 68202

RESULT 4
T7M24
LOCUS      T7M24
DEFINITION Arabidopsis thaliana BAC T7M24.
ACCESSION  AF077408
VERSION     AF077408.1 GI:3319359
KEYWORDS
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 74475)
Harmon,G., Langston,Y., Stoneking,T., Drone,K. and Ames,M.
The sequence of Arabidopsis thaliana T7M24
Unpublished
2 (bases 1 to 74475)
Washington University Genome Sequencing Center.
The Arabidopsis thaliana Genome Sequencing Project
Unpublished
3 (bases 1 to 74475)
Waterston,R.
Direct Submission
Submitted (10-JUL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
COMMENT

```

e-mail: rwilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by a sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is T5H22 ; 3' cosmid is T25H8 . Actual start of this clone is unknown ; actual end is at 74475 of T7M24.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

Location/Qualifiers
1. 74475
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="unknown"
/clone="T7M24"
4936. 7830
/gene="T7M24.7"
4936. 7830
/gene="T7M24.7"
/note="contains similarity to reverse transcriptase (Pfam: rvt_hmm, score 19.29)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC26250.1"
/db_xref="GI:3319362"
/translation="MVLEKNCYVYPAINKNLIISVCLDMEGFHFISIKNKKCSFDRDD
MEYGSAPLDNGLHVNLQSNMFIYNIIRTKFKSNDLNPFLWHCKRLGHINEKHQKHS
GLNLFDESYETCSCLGKMTKAPFTGHSERASDLGLIHTDVGPMSTARGNYQ
YFITTDDFSRYGYVLMKHSFENFQNEVQNFQKGIKALRSRGGYLSQV
FSDHRECGIVSQLTPTGTPQWNGVSRNRNTLLDMVRMMSTDLPSFPGVYALTS
AFMLNRCFSKVEKTPYEITWKVPLNLSFKIWKGCESYAKRLITDKLGPCKCYFVG
YPKETGYVYHPTDKVFNVGAFLEEFISKGTSCSKVLEEVREPCQDVPYTSQE
EHOLDLRVVEPILVEPEVRSRSHREDRFRDWMDDHALLIESDPTSEALM
GPDCKWLEAKSEMSSQNKWTLVLDLPGVKPIEKWIIFKMDMDGNTQIYKQ
LVAKGYKQVHGIDYDTSYVAMLKSIIRLLATAAHYDEIYQMDVKTAFNLNLEEH
VYMTQEGTVPAAKCKLHRSYGLQASRWNLRNEAIKEPDFRNNEEPCVY
KKTSGSAVFLVYDDILLGNDIPLQSVKTLGSCFSCMDMGEAAVILGTRVYD
RLNKIIGLSQDVIYDKVLRFNHMDSKGFIEMSHGITLSKTCQSPSTHDERMSKIP
YASAGSIWYAMLYTRPVACALMSYRSQSDPGSHWIVVRNIYKLRRTKDKFLVY
CGSEELVSGYTDASFDQKDFRSQSGFFCLNGAGIQAQKEPKSKHQRRI
AASEAAKEVWIRKEFITELGVVPSGIPDLYCDNNGAIAQAQKEPKSKHQRRI
HLIREIIDRGVWKISRVSTADVADHFTKPLPQPKHSHHTAIGIRFIKM"
15079. 19726
/note="pseudogene"
complement(21094..21816)
/note="pseudogene"
41914. 45895
/gene="T7M24.5"
join(41914..42178,42523..42709,44317..44441,44538..45141,

60988 CTCGACCGCTTCGCTCAAGTGTACTGACCTCTTCCACCCTCAAGCAACAAAACCTCGA 61047
754 AAAAAGAAAAAAGGTCACAAACGACAGTTCCTCAACCGGTTTACTCGACCGGTAC 813
61048 TTTCCAAAACCAACAAAATCACTCGCGCGTTT-ACTCAATAACATTAATCTTGAACCAATTT 61106
814 GCTGGTTTAGATGTTGTTTGGTTTCTGCTATTAATAACAATTAATCTTCTTGAAG 873
61107 ACTCTGTTAGTTTATGTTTGGTTTCTGCTGGAATTAACGTTTATCTTCTTGAAG 61166
784 TTTGCTGTTTGTAGGTTTCTCATGAGTAACTACAGTGAAGAAATCTCTATGAGACCTT 933
61167 TTTTCATCTATTTTTCAGGTTTCTCATGAGTAACTACAGTGAAGAAATCTCTATGAGTCTT 61226
934 GATTATAATGTGATGAAGCTTAAGTTCCTGGTTCCTACTAGACCGGAGTGAGACCAATCTTT 993
61227 GATTATAATGTAGATGAAGCTGAGTCTTGGTTCCTACTAGACCAAGAGAGAAACAAGTTT 61286
994 TACGAGAGCTATAGGATGAATTTGAACGCTCTGAGCTCGACGTGACGTAATCAAAAGAGAGCT 1053
61287 TACGAAAGAGCTTCAAAGAGAGATTTGAAACGCTCTGAGCTCGACGTAAACAGAGAGAGCT 61346
1054 GAAATCGCTAGAGAAAGAGAGCGCATGTCGAGTAGATATGAGCTGATTGATGAGGATATC 1113
61347 AAATCCCGCAGAGAAAGAGAGCGCATGTCAGTAGATATGAGTGTGATGAGGACATC 61406
1114 AAACTGAGTATGAGCCAGAGTCAATGGCGCAAGAGAGACGAAGCTTACTGAAACAAATCCGAC 1173
61407 GAAACTGAGTATGAGCCAGAGTCAATGGCGCATGAGAGAGAGCTGCTGAAACAAACCCGAC 61466
1174 GAGGTTACGTTGAGGAGTATATCAGATTTCTTTGAGATGAATGATCTTCTGGGGAACGAGG 1233
61467 GAGGTTACGTTGAGGAGTATATCAGATTTCTTTGAGATGAACGACTTCTGGGGGACGAGA 61526
1234 TATCCCTGATATGAGACTTTAGCCCGCTGGGTTACTGAGAGACGTCGAGCATCTGTTTC 1293
61527 TATCCCTGATATGAGACTTTAGCCCGCTGGGTTACTGAGAGACGTCGAGCATCTGTTTC 61586
1294 GAGAAGTGTCACTG-----ATAAGGAGGAGACA 1322
61587 GAGAAGTGTCACTGAGAGACACTCATGCTTACCCCTTACGCTACTATTAAGAGAGAGG 61646
1323 ATCGAGTTCTTTCACACATGCAAGTGGAATGATGAGGACTCAGACTTTTGAGCTG 1382
61647 ATCGAGTTCTTTCACACTGCAAGTGGAATGATGATGATGATGATGATGATGATGATGATG 61706
1383 GATACCATGGGTTAGGCTTCTTGAGCTTCTTAGTGATCAACAGCGGTACAGATTAG 1442
61707 GATACCATGGGTTAGGCTTCTTGAGCTTCTTAGTGATCAACAGCGTTATCAGCTGTGCG 61766
1443 ATCAAGAAATTTGAAGAACTGTTTGGTTTCCCTTAGTGGAAGGAAACCAACCCAGGTTT 1502
61767 ATCAAGAAATTTGAAGAACTGTTTGGTTTCCCAAGTGGAATGGAACCAACCCAGGTTT 61826
1503 GACAGGAGAGCTTGAAGATTTGCGGCTACTATTTGGGAACAATCTACCGTAAACTCG 1562
61827 GAAAGAAAGAGCTTGAAGATTTGCGGCTACTATTTGGGAACAATCTACCGTAAACTCT 61886
1563 ACGCGGTCCAAGAGCAACAAATCCGAGTCTGTTGATGATGATGATGATGATGATGATGATG 1622
61887 GCGCGGTCCAAGAGCAACAAATCAGAGTCTGTTGATGATGATGATGATGATGATGATGATG 61946
1623 GCAATGTTTTTACTCCAGGAGTCTACAGGACCGCTGCTTAAACAGACATGAAGATG 1682
61947 GCAATGTTTTTACTCCAGAGAGTCTACAGGACCGCTGTTTAAACAGGACATGGAGATG 62006
1683 ATAGATTACGCGCTTATAGGATTTCTCCGCTTACAAAAGGAAGAAATGTCCTGAGAGGA 1742
62007 ATAGATTCCGCACTTACAGGAAATTTCTCCGCTTACAAAAGGAAGAAATGTCCTGAGAGG 62066
1743 GATCTTAAGACTCACACAGTAATGCTCTGTTGATCCATCTGTTGGGTACATGAAG 1802

Db 62067 AATCTCAATGACGGCCACCAGTAATGCTCTGTTGATCCACCTGTTGATACAGGAAG 62126
Qy 1803 TGGCGCTGACAAACGGCAAGAGGTAAGAGAGACACTATGCTGGGTGGCCCTTGTG 1862
Db 62127 TGGCGCTTGAACAACGGCAAGAGGTAAGAGAGACACTATGCTGGGTGGCCCTTGTG 62186
Qy 1863 ACGCAATTTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGTTAGCACCGGAGATG 1922
Db 62187 ACGCAATTTCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGTTAGCACCGGAGATG 62246
Qy 1923 ATGCACTTGGATCACTTGGCGGATGTGAGTTCTCTGAGTTTGCATGTTGGGACATTT 1982
Db 62247 ATGCACTTGGATCACTTGGCGGATGTGAGTTCTCTGAGTTTGCATGTTGGGACATTT 62306
Qy 1983 CACCGCTACAGGTTTCGAGCATTCATCATTAAGATAGCAATCTTCTTCTCCCTGCAAT 2042
Db 62307 CACCGCTACAGGTTTCGAGCATTCATCATTAAGATAGCAATCTTCTTCTCCCTGCAAT 62366
Qy 2043 TACGCTACTAGGATTTCTCGAGGGCAGCAACATTAAGATAGCAATCTTCTTCTCCCTGCAAT 2102
Db 62367 GACGCTAATAGGATTTCTCGAGGGGAGCAATTAAGATAGCAATCTTCTTCTCCCTGCAAT 62426
Qy 2103 TATTTGAGGCGAGTCCGCAACTGAGGAGATTAAGATAGCAATCTTCTTCTCCCTGCAAT 2162
Db 62427 TATTTGAGGCGAGTCCGCAACTGAGGAGATTAAGATAGCAATCTTCTTCTCCCTGCAAT 62486
Qy 2163 GATGTTGATGAGACATATGATATAGATGAGCGGAGTTTGCACAGGAGCATGTATCTTTC 2222
Db 62487 GATGTTGATGAGACATATGATATAGATGAGCGGAGTTTGCACAGGAGCATGTATCTTTC 62546
Qy 2223 AGTGAGCATATATCTTCAGCGAGGAGAAAGCAAGAGTTTGCAGGAGCTTCACAGGAGCAAC 2282
Db 62547 AGTGAGCATATATCTTCAGCGAGGAGAAATCAAGAGTTTGCAGGAGCTTCACAGGAGCAAC 62606
Qy 2283 AGCAAGCTGAGAGTGGTTCAGAGAAACAGGATTAAGTTCGCAAGTGGCTTCAGGCT 2342
Db 62607 AGTAAGCTGAGAGTGGTTCAGAGAAACAGGATTAAGTTCGCAAGTGGCTTCAGGCT 62666
Qy 2343 ATCAAGTCTTGAAGCAACAGATCAGTGTCTCTCTTCTTCTCACTCACTATTCGCAATGA 2402
Db 62667 ATCAAGTCTTGAAGCAACAGATCAGTGTCTCTCTTCTTCTCACTCACTATTCGCAATGA 62726
Qy 2403 CAGCTCCCTCAGGACATGCTTCGAGGAGATATGAGCGGCTTAAGCAAGTTCGCGGAGG 2444
Db 62727 CAGCTCCCTCAGGACATGCTTCGAGGAGATATGAGCGGCTTAAGCAAGTTCGCGGAGG 62786
Qy 2445 ---GAGCTAGAGAGAGAGATTTCTGATGTCCTCTCGAGGCAATTCATCATTCGAGCT 2501
Db 62787 CCGTGGCTAGTGAAGAGAGATTTCCGATGTCCTCTCGAGGCAATTCATCATTCGAGCT 62846
Qy 2502 CGTGATCTAGGAGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2561
Db 62847 CGTGATCTAGGAG 62906
Qy 2562 CTTCTGAGCTCTGATGTTTACCGACCGGCTGCTGGCGCAATTAAGAGAGAGAGAGAGAG 2621
Db 62907 CTTCTGAGCTCTGATGTTTACCGACCGGCTGCTGGCGCAATTAAGAGAGAGAGAGAG 62966
Qy 2622 GAGTATCTCAGAGCGGCTGCTGGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2681
Db 62967 GAGTATCTCAGAGCGGCTGCTGGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 63026
Qy 2682 GAGTATCTCAGAGCGGCTGCTGGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2741
Db 63027 GAGTATCTCAGAGCGGCTGCTGGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 63086
Qy 2742 GACTACCAACTCTGCTTCAATTTTCGACTGAGGTAAGCGCTCAGCTTCACCATATATAT 2801
Db 63087 GACGACCAACTCTGCTTCAATTTTCGACTGAGGTAAGCGCTCAGCTTCACCATATATAT 63146
Qy 2802 ATCATCTCTGAGATTTGTTCTTTATTTTTCAGTGAATGAGATTTGCTCTGAGTACTC 2861
Db 63147 ATCATCTCTGAGATTTGTTCTTTATTTTTCAGTGAATGAGATTTGCTCTGAGTACTC 63206

QY	2862	TCCTCCAAAGTTTATTCACACAGTGCAGTGTGTGATTTAAAGTTTGGGGAGGGCTCAGGAA	2921	from 51105 to 53053 element 17 resides from 53054 to 5499 element 18 resides from 54998 to 56946 element 19 resides from 56947 to 58860 element 20 resides from 58861 to 60772 element 21 resides from 60773 to 62663 element 22 resides from 62664 to 64578 partial element from 64579 to 65373		
Db	63207	TTTTCCAAAGTTTATTCACACAGTGCAGTGTGTGATTTAAAGTTTGGGGAGGGCTCAGGAA	63266			
QY	2922	--GTATGTTGCATTTATATATATTTTAAAGTCTGCATTCATCATCAAGGATGAAACCAA	2979	48761..48946		
Db	63267	GTGTGTGTTGCATTTTATATATCTTGTAGTCTGCATTCATCATCAAGGATGAAACCAA	63326	/note=function=centromeric repeat similar to fragment of mi167, GenBank accession number X94367"		
QY	2980	AAAAAAATTTAAAAATTTTCAGAAATGATTTTCACAAAAAAGAGTGTTCATGTAGTTCAT	3039	49937..52753		
Db	63327	AAAAATTTGAAAAATTTTCAGAAATGATTTTCAC--AAAAATGATGTTCATGTAGTTCAT	63384	/gene="AT4g03800"		
QY	3040	TACATTTAGGATCAAGTCTCAGAGTCTTTTCATTTAGGATTTGATATGATAGGAGATA	3099	complement (join(49937..50481,50527..50943,51178..51438))		
Db	63385	TACATTTAGGATCGAGTCTAGAGTCTTTTCATTTAGGATTTGATATGATAGGAGATA	63444	/gene="AT4g03800"		
QY	3100	ATGATGAGATAGCTTGTAAGCA	3122	/note="similarity to"		
Db	63445	ATTATGATATAGCTTTGTAAGCA	63467	/codon_start=1		
RESULT 5						
ATCHRIV10		ATCHRIV10	196296 bp	DNA	linear	PLN 16-MAR-2000
LOCUS		Arabidopsis thaliana		DNA chromosome 4,	contig fragment No. 10.	
DEFINITION		Arabidopsis thaliana				
ACCESSION		AL161498				
VERSION		AL161498.2	GI:7267134			
KEYWORDS		Arabidopsis thaliana (thale cress)				
SOURCE		Arabidopsis thaliana				
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustoids II; Brassicales; Brassicaceae; Arabidopsi				
REFERENCE		1 (bases 52654 to 127128: 151832 to 196296)				
AUTHORS		Lamar,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 196296)				
AUTHORS		EU Arabidopsis sequencing,project.				
TITLE		Direct Submission				
JOURNAL		Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk				
COMMENT		Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV9 at the 5' end and an overlap with ATCHRIV11 at the 3' end.				
FEATURES		Location/Qualifiers				
source		1..196296				
		/organism="Arabidopsis thaliana"				
		/mol_type="genomic DNA"				
		/variety="Columbia"				
		/db_xref="taxon:3702"				
		/chromosome="4"				
repeat_unit		5001..48473				
		/note=function=unclassified 22 tandemly arranged copies of a 1950-bp element each element is composed of a 1890-bp core flanked by 2 31-bp direct repeats element 1 resides from 21901 to 23849 element 2 resides from 23850 to 25798 element 3 resides from 25799 to 27743 element 4 resides from 27744 to 29665 element 5 resides from 29666 to 31621 element 6 resides from 31622 to 33582 element 7 resides from 33583 to 35520 element 8 resides from 35521 to 37468 element 9 resides from 37469 to 39416 element 10 resides from 39417 to 41364 element 11 resides from 41365 to 43313 element 12 resides from 43314 to 45261 element 13 resides from 45262 to 47209 element 14 resides from 47210 to 49155 element 15 resides from 49156 to 51104 element 16 resides from 51105 to 53053 element 17 resides from 53054 to 54998 element 18 resides from 54999 to 56946 element 19 resides from 56947 to 58860 element 20 resides from 58861 to 60772 element 21 resides from 60773 to 62663 element 22 resides from 62664 to 64578 partial element from 64579 to 65373				
		/note="function=putative_transposon similar to TNP2 of A. majus similar to F1X3, GenBank accession number AC008266 similar to T4E14, GenBank accession number AC005171"				
		52654..52853				
		/note="overlap with T5H22, GenBank accession number AF096372, from position 69553 to 69752"				
		53634..54512				
		/note=function=transposon similar to TNP2 of A. majus similar to F1X3, GenBank accession number AC008266 similar to T4E14, GenBank accession number AC005171"				
		57589..60483				
		/gene="AT4g03810"				
		57589..60483				
		/gene="AT4g03810"				
		57589..60483				
		/note="strong similarity to"				

QY	1054	GAAATCGCTAGAGGAAAGAGGGGATGTCGAGTAGATATAGCTGATTTGATGAGGATATC	1113	2103	TATTTCCAGGCGAGTCGCGCAACTGAGGAGATTAGTCAACGGAAGGAGCTACATAGAA	2162	
Db	114000	AAATCCCGAGAGAAAGAGCGATGTCAGTAGATATGATGATGATGATGAGGACATC	114059	Db	115080	TAITTTGAGGCGAGTCGCGCAACCGAAGAGATCAGTCAACAGAAGGAGCTATAAGGAA	115139
QY	1114	AAAACTGAGTATGAGCCAGAGTCATGCGGCAAGAGACGAGCTACTGAACAAATCCGAC	1173	QY	2163	GATGTTGATGAGACATATGATATAGATGAGCGGAGTTTGAACAGGAGCATGATCATTTTC	2222
Db	114060	GAAACTGAGTATGAGCCAGAGTCATGCGCATGAGACGAGCTGCTGAACAAACCCGAC	114119	Db	115140	GATGTTGATGTCACAGTTGAAATAGATGAGGTGGAGTTTGAACAGGAGCATGATCATTTTC	115199
QY	1174	GAGTTACAGTGGAGGAGTATACAGATCTTTTCAGATGAATGATCTCTCGGGAACGAGG	1233	QY	2223	AGTGACATATACCTCCAGCGAGGAAAGCAAGAGTTTTCAGCGGAAGCTCACAGGAACAAC	2282
Db	114120	GAGGTTACGGTGGAGGAGTATACAGATCTTTTCAGATGAACGACTCTCTCGGGAACGAGA	114179	Db	115200	AGTGACATATACCTCCAGCGAGGAAATCAAGAGTTTTCAGCGGAAGCTCAAAGAACAAC	115259
QY	1234	TATCCCTCATATGAGACTTTAGCCAGTTGGGGTTACTTGGAGGACGTGACAGATCTGTTTC	1293	QY	2283	AGCAAGCTGAGAAAGTGGTCAAGAAACAGGATAAGTTACTCGCCAAAGTCCCTCAGGGCT	2342
Db	114180	TATCCCTCATATGAGACTTTAGCCAGTTGGGGTTACTTGGAGGACGTGACAGATCTGTTTC	114239	Db	115260	AGTAACTGACAGAGTGGTCAAGAACTGGATAAGTTACTCGCCAAAGTCCCTCAGGGCT	115319
QY	1294	GAGAGTGTCACTGTG-----ATAAGGAGGAGCA	1322	QY	2343	ATCAAGTTTCTGAAGACAAAGATCAGCTGCTCTCTTCCACTACACTATTCGCGCAATGA	2402
Db	114240	GAGAGTGTCACTGTGAGACACTCATGTCTTACCCCTTACGCTACTATATAAGAGGAGG	114299	Db	115320	ATCAAGTTTCTGAAGACAAAGCTTAGCTGCTCTCTTCCACTACACTATTCGCGCAATGA	115379
QY	1323	ATCGAGTTCTTTCCACTGCAAGTGAAGATGATAGGACTCACAGACTTTGAGCTG	1382	QY	2403	CAGCTCCCTCAGACATGCTCTCGAGGAGATATGACGCGCC-----2444	2444
Db	114300	ATCGAGTTCTTTCCACTGCAAGTGAAGATGATAGGACTCACAGACTTTGAGCTG	114359	Db	115380	CAGCTCCCTCAGACATGCTCTCGAGGAGATATGACGCGCTTAAGCCAAGTCGCGCTAGG	115439
QY	1383	GATACCATGGGTTAGGCTTTTACAGTTCTTAGTGGATGAACAGCGGTACAGATTTAG	1442	QY	2445	---GAGCCTAGAGAGCAAGAATTTCTGCATGTCCTCTCGAGGACTTATCATTTTCGAGCCT	2501
Db	114360	GATACCATGGGTTAGGCTTTTACAGTTCTTAGTGGATGAACAGCGGTACAGATTTAG	114419	Db	115440	CCTGAGCCTTAGTGAAGGAGATTCGCCATGTCCTCTCGAGGACTTATCATTTTCGAGCCT	115499
QY	1443	ATCAAGAAATTTGGAAGAACTGTTGGTTCCCTAGTGAAGAGGAAACCAACCCAGGTTT	1502	QY	2502	CGTGAATCTTAGGAAGATAGAGGAACGACACTCATCTGATCTAGCAGCAGGAGCAGACGA	2561
Db	114420	ATCAAGAAATTTGGAAGAACTGTTGGTTCCCTAGTGAAGAGGAAACCAACCCAGGTTT	114479	Db	115500	CGTGAATCTTAGGAAGAGGAGGACCACTCATCTCAATCTAGCAGCAGGAGCAACAGGA	115559
QY	1503	GACAGGAAAGACTTAAAGATTTTGGGCTACTATTGGGAACAATCTACCGCTAAACTCG	1562	QY	2562	CTTCTGAGACTCTGCTAGTTTACGCGACCGCGGTGCTCGCGCCCAATGAAGAAGAGAGGTC	2621
Db	114480	GAAAGAAAGACTTAAAGATTTTGGGCTACTATTGGGAAACAATGTACCACTAAACTCT	114539	Db	115560	CTTCTGAGACTCTGCTAGTTTACGCGACCGCGGTGCTCGCGCCCAATGAAGAAGAGAGGTC	115619
QY	1563	ACGCGTCCAGAGCAACCAATCCGAGTCTGTGATTCGCTACTTTTCAGCGCTCGGTT	1622	QY	2622	GAGTATCTCTCAGAGCGGTGCTGCGGCCACAGAGCTGATGAGATCGAGTACCCATGCT	2681
Db	114540	GCGCGTCCAGAGCAACCAATCCGAGTCTGTGATTCGCTACTTTTCAGCGTTCAATT	114599	Db	115620	GAGTATCTCTCAGAGCGGTGCTGCGGCCACAGAGCTGATGAGATCGAGTACCCATGCT	115679
QY	1623	GCCAAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTAAACAGACATGAAGATG	1682	QY	2682	GGAGTCTGATACGGAACATGCGGTTGCTGTATGCGCTTGGGAGCAATCACAGGAGCCATT	2741
Db	114600	GCCAAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTAAACAGGACATGAAGATG	114659	Db	115680	GGAGTCTGATACGGAACATGCGGTTGCTGTATGCGCTTGGGAGCAATCACAGGAGCCATT	115739
QY	1683	ATGATTTACGCGTTATAGGATTTCCGCTTACAAAGGAAGAAATCTCTGAGAGGA	1742	QY	2742	GACTACCAACTTCTGTTTATTTATTCGACTGAGGTAAAGCGCTCCTTCAACCAATATATAT	2801
Db	114660	ATGATTTACGCGTTATAGGATTTCCGCTTACAAAGGAAGAAATCTCTGAGAGGC	114719	Db	115740	GACCAACCACTCCGTTTCAATCTTCGACTGAGGTAAAGCGCTCCTTCAACCAATATATAT	115799
QY	1743	GATCTTAAAGACTCACACAGTAATGCTCTGATCCATCTGTGTGGGTACATGAAG	1802	QY	2802	ATCATCTCTGATTTCTTCTTATTTTGTTCAGTGAATTTGATTTGCTCTGAGTACTC	2861
Db	114720	AATCTCAATGACCGGCCACAGTAATGCTCTGTTGATCCACTGTGTGATACAGGAAG	114779	Db	115800	ATCATCTCTGATTTCTTCTTATTTTGTTCAGTGAATTTGATTTGCTCTGAGTACTC	115859
QY	1803	TGGCGCTGACAAACGCGCAAGAAAGGTAAAGAGGACACTATGCGTGGGTGCGTTGTG	1862	QY	2862	TCCTTCCAAAGTTTATTCACACAGTGGACTGTGTGATTTAAGTTTGGGGAGGCGCTCAGAA	2921
Db	114780	TGGCGCTGACAAACGCGCAAGAAAGGTAAAGAGGACACTATGCGTGGGTGCGTTGTG	114839	Db	115860	TTTTTCCAAAGTTTATTCACACAGTGGACTGTGTGATTTAAGTTTGGGGAGGCGCTCAGAA	115919
QY	1863	ACGCCAATTCGAAAGTTTGTGAGTTCGCTCAAGGAGTGGGTTAGCACCAGGAATG	1922	QY	2922	--GTAATGTCATTTATATATTTTAACTGTCATTTTCACTTAAGGCATAGAAAACCA	2979
Db	114840	ACGCCAATTCGAAAGTTTGTGAGTTCGCTCAAGGAGTGGGTTAGCACCAGGAATG	114899	Db	115920	GTGTGTGTCATTTTATATTTTGTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG	115979
QY	1923	ATGGACTTTGATCATTTTGGCCCGATGAGTTCTTGAGTATGACATGTTTGGGACTTC	114959	QY	2980	AAAAAATTTAAAAATTTTCAGAAATGATTTTCAAAAAAAGAGTCTTTCATGTAGTTGCAT	3039
Db	114900	ATGGACTTTGATCATTTTGGCCCGATGAGTTCTTGAGTATGACATGTTTGGGACTTC	115019	Db	115980	AAAAAATTTAAAAATTTTCAGAAATGATTTTCAAAAAAAGAGTCTTTCATGTAGTTGC	

RESULT 6
T15F17/c
LOCUS T15F17 68352 bp DNA linear PLN 26-JUL-2000
DEFINITION Arabidopsis thaliana BAC T15F17.
ACCESSION AF262042
VERSION AF262042.1 GI:7682788
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 68352)
Washington University Genome Sequencing Center.
The A. thaliana Genome Sequencing Project
JOURNAL Unpublished
AUTHORS Wilson, R.
REFERENCE 2 (bases 1 to 68352)
Direct Submission
Submitted (01-MAY-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: rwilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one m13 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

FEATURES

source
1. 68352
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="v"
/map="unknown"
/clone="T15F17"
13450..13599
/gene="T15F17.j"
13450..13599
/gene="T15F17.j"
/codon_start=1
/evidence=not experimental
/product="Hypothetical protein T15F17.j"
/protein_id="AAF67378.1"
/db_xref="GI:7682798"
/translation="MTALSYYEKKRRERANQEKAMISDKVMSGTLEDRKLRAVE
KAKQP"
complement (16795..23597)
/gene="T15F17.i"
complement (join(16795..18896,20555..20940,20995..22337,

23343..23597))
/gene="T15F17.i"
/codon_start=1
/evidence=not experimental
/product="Hypothetical protein T15F17.i"
/protein_id="AAF67377.1"
/db_xref="GI:7682797"
/translation="MAAILPVKGSTKTEWNTVTHIESTSRPSCIVLLSHGETDQR
SVRPKGSTDNOLSPKPKMDTIDVGPSPKADGRKISWKNKDVBEWELVENLAQS
DGYNEODRTVGTADSDDKRKDIKALNDKMLLSQKHVHFLVDDRYQVQDG
EGNOLSEVYINNNGYKNNPNTNNPNLSYRNTVANPDQVYPTQQQGNKPF
VLNQDFVPKQFQNYHQPFPQPAALQKQPAALDAEMKEMKQVLLQQAQSSMEI
AKLSLHKLDCSYNNLNKVALEANTKRYLEGQASTSTLTKVTLGPKGSIQNKY
ATAHAITICHRELPTOPVDFITGNDNDVOGEASTQVEYSIVFNHSGRHLTQST
SEEKAALIERWVKRFKPTPLPSRALPMTFSKAWNERKYSVAQKQLEAEVMPLEVL
NLAPDHKVRNLILRIKLYHDSDESATLSAAKRIQVEKLEDPGSTLPSCISG
EFAPSDCLDLGASVSLPMSVARSRRKPFGLKOLPVMINRVEVPTDFVLDMEVE
HKDPLILGRPFLASVGAVIDVREGNIGNLGKHIKLFNINRTPGQSTEDGTGRNDR
AISGEGYETREVKELKRSNKQDETIKLAHTTEELRSFTMSNYSGFSDPDYNDRE
AESWARSREQHVLOSFRDEKRSARPNORRAEIAARGKMSRYRVEDIETEYS
PEWRRETKLLNKPDEVTVEYIRPFEMDNDFWGTLYCYETLAQLGLEDVQHLFKKC
HLETLMSYPYPAKKEETIEFLTIQVMEYALTDIFELDIMGLELTVLNEQYQLSI
KKLEELFGPFGSGTKPRFREELKDLMTIENNLPLNLSKSNQIRSHVICYFQRS
VANVFSRESTGVFNFDIEMIDYALTILGRTKGNVLRGLDNNAPQVWPLLTFFCG
YRKWALNTGKKVKGALCVGGVTPILQCGCVLPKEGLAPRMWDLHLRCEFLFD
MAGDFHYRFDNSSI KIANILLPCIDATRIEGRNIDFKPALBELYEGSPPTTEISH
TEGATTKIDITNIDEAEDTSMYHSEHPPARESRSLSEAHNRNKKLOKCMKOD
KLLAKHRAIKFMKNKLSYSSNTAIPQGHLPQDMPSPRYDAPEPSRRHLPQCEIP
LVHARSSEFPESGRRKRTTLTRSSRGLRSLRSLRDRGAGRRRREVEYPOQSA
GRHRADEVYPPAGADTEQGSLSITSYHSSTVSASLHHCIIPISTIVICFFIFFISV
IGVLSLTFOVYSHGLCLSLGEGSGVCCIVHIFESAFI"
complement (24606..24944)
/gene="T15F17.h"
complement (24606..24944)
/gene="T15F17.h"
/codon_start=1
/evidence=not experimental
/product="Hypothetical protein T15F17.h"
/protein_id="AAF67376.1"
/db_xref="GI:7682796"
/translation="MAEEQNQNGPANIGVDAPDRHQSKGIEPPAIQNNNFEIKSG
LSMIQGNKPHGLMEDPLDHLDEFDRHYNLKKINGVSDGFKLCLFPFSLGNKAHIV
EKNCNARVPS"
29265..30290
/gene="T15F17.k"
29265..30290
/gene="T15F17.k"
/codon_start=1
/evidence=not experimental
/product="Hypothetical protein T15F17.k"
/protein_id="AAF67379.1"
/db_xref="GI:7682799"
/translation="MSKLTNMYAALLAFGDNLYQWALDTKIELSKGLAEICVEGNE
KTERSKYQAIIRHIAELSKNYLTVEDPLELWELKNRYDHQRTIOLPKAQHDWL
NLRIQDYKSEVYNSGLFKIVSLILCGKVTENDMLEKTFTSTFANNVLLQOYRAK
GFTVTSLASCILLAEKNELLIMNSALFPQGSTAVPEANRAEMAKAPNPQATKSN
VVRGNPHGCHGRGGRGNGFYCQGNHYGCRGNGYRSGRGRGIRKPKGKKS
VCYRCGMDHWAKTCTSKHLIEAQEMIKQKGPANLVHLDEGDFDEKDLMDYE
TSDIILGE"
30441..35665
/gene="T15F17.l"
join(30441..31166,31218..33390,33445..33789,35484..35665)
/gene="T15F17.l"
/codon_start=1
/evidence=not experimental
/product="Hypothetical protein T15F17.l"
/protein_id="AAF67380.1"
/db_xref="GI:7682800"
/translation="MAEDKILIVDSGSSHTILDRKRYFINLTNNANISTAGIASL
LEGYQAHVLLPNTGTHLEISDALYSPSSKSLLSFKDLNGFHVHTGEGNREFLHI
TEIAQGHKRVLETTPALSTGLYHTKINMINMEANLANKFEIREFTHDRLDGPHNMW
RKLMISSKGTHTKEKRVIPKLTCAACQAKLIIRPSAKYNKKTINFLERTQGDICG
PIHPGCGTFRFMVILIDASTRWSHILRLRAHFDPDFPKLTIRLDNAGEFTSQAFNDYCM


```

SMGVSEHPVAHVHTQNGLAESFIKRIQLIARPLLMBRSLPVAAGHVAHLHSELIRI
RPSSEHKISQLLTHGHPDLSHLRTGCAVYVPIAPQTKMGPORMGIVYGFQSP
TIIKYLEPTGLDIFKARYADHLPSESEFPALGGKTNKLGKKEIKELIWNQISLWQDPQ
TLMCDSEVHKI IHLQKLANELPSPADPKRVKSYIPACNAQVRIDILQKGNLVQATPS
NPKKRGRPVSGDKDKPRKTKGAMTEVKEITDMAAASPKEPSIEIRDAEAGHEGPI
DNBISLNYTMSGIKMNRKVDVDEKAYEVAVEINEDHEPTSIIECTQSDMLKWKBE
AIDVELGSLKRDVFGPILRTPSNIKPVGHKWVVRKRNEKNIIVRHKARLVAQGFQ
RPGVDYBETSPVDATITFRYLISLAIRENLDRLMDVVTAYLYGPILDNELYMLPBG
TELKGNKNSRDQYCRIMNRSYLGKSGRMWYNRLSEYLVKRGYKNDPISPCIFIK
KFASKGVFIITAVYVDDILNLTGEIATQVEYLKKBEMKDLGKTKFGLQLGLYVOK
GILVHORAYETILKRNMDKAHLPTSPMVQVSLGSDPFPKDDDEEILGPEMPYL
SAJCAKGVYVSHTEPDTICFANLLSRFSSCPKRWHEGIIKHLRYLOQTIDFGYLSDP
HNGSQYVYVTHGGTALSWFSMKOTIAATSSNHAELIAMEHASRECVMWLSMTQHIR
ADCGWEDKEFTIIEYDNNACIAQLKSGYIKGDETKHLLPKFFTHDLQKDVCEPDA
AVCEPDAAAFPDATAFSDPANLPTDLADLPDAADLPDAADLPDAADLPDAADLPDA
37905..38945
/gene="T15F17.m"
CDS
/gene="T15F17.m"
/gene="T15F17.m"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein T15F17.m"
/protein_id="AAF67381.1"
/db_xref="GI:7682801"
/translation="MADVVDROEPTNIGADFPFHHNRHGIYVPPVQNNFKIKSS
LIAMVOGNKFGFLPWEDPLDHLDFERLCLDKINGVSEDFKLRLPFSLGDKAHLW
EKLTPQSGITWDDCKKVFLEKFPNSRTARLWNEISGFTQKQNESFCEANERFKGYO
TKPHGGFKASLLSTLYRGLPKIRMLLDLTASNENFLNKDVEDGWELVENLAQSDGN
YNEVDRIPTSTESDDKHREIKALHEKIDKLQVQKHVSHASEDBVFQIQEEND
QDABTSYVQNOGGYKNGYNNRPNLTVRSTNVANPQNVYVPOQHSRTKPNRLFTT
KAKGSFLILIGP"
complement(42508..43164)
/gene="T15F17.g"
CDS
complement(42508..43164)
/gene="T15F17.g"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein T15F17.g"
/protein_id="AAF67375.1"
/db_xref="GI:7682795"
/translation="WVDDDFKFRFEHSTDRRANILLPSEVTRIIEGNDIRFPEVGR
LYENAPLDDDEFELEAASDGMEDRAVEFDTSVMYFGEHVPFAPRQSKSLTKPHKN
SKLQKWKHKDKLIAKCPKLOTDKLSGSSSTTAIPQVQPPLEMPRRFNAPARPELS
EORVPHVQARHSFSPSREHRRKRLTLTSSRSRLIHSRRSLDCCAGRSRRREVEYP
"
complement(46220..47737)
/gene="T15F17.f"
CDS
complement(46220..47737)
/gene="T15F17.f"
/codon_start=1
/evidence=not_experimental
/product="Hypothetical protein T15F17.f"
/protein_id="AAF67374.1"
/db_xref="GI:7682794"
/translation="MTLMSYPVPAYKEETPEFLSTLQVMEYALDTELDITMGLFL
TSPVNEQRYELSIKLEELFGPSPGKTRPDRBELKDLWATIGNNLPLNSARSKN
"
Query Match 63.2%; Score 1973; DB 8; Length 68352;
Best Local Similarity 82.4%; Pred. No. 0;
Matches 2498; Conservative 0; Mismatches 395; Indels 138; Gaps 15;

Qy 163 CTGAGCTCGCGCTTCTACCGCCTCTCCATCGTACCGCCTGCTCCCTCTCTCCAAAGG 222
Db 48823 CACTCGATCGCGCTTCTACCGCCTCTCCATCATCGTTGCTACTCACTCTCTCCAAAG 48764

Qy 223 AAACAACTCGAGCTCTCCATTCTACTCGAGCTCTACCAACCAAGCGGCTTCACCAC 282
Db 48763 AAACAACTCGAGCTCTCCATTCTACTCTCACTCGAGCTCTC----- 48725

Qy 283 TTCTAGCTCTTAACCACTCGAGCCCTTCACCATCAACCAATCAAACTGTTTCTCTCC 342
Db 48724 -----CACCATCAACCAACCAAAATGGA-TTCTCTCTCC 48694
```

```

Qy 343 ATTAAGCTTGACANACTCGACCGCTGAAACACTTATCACTTCAAGCTCCTCATCTCTTC 402
Db 48693 ATTACAGCTTTC-----ACTCGACCGCGCAACATTTATCACTTTAAGCTCTCATCTCTCC 48638

Qy 403 ATCGTTTCCAAACCGCTGCTCTCATCCGCCACGAAAGCTTGTTCATCACCTCTCACTCAT 462
Db 48637 ATCGTTTCCAAACCGCTGCTCTCAT--CCACGNAAGCTCGTCATCTCCTTTCACTCAA 48580

Qy 463 CACCAGTTCACTCGATTTCAGCAACCAAACTCGACTCGTCTCTCTTGGCCACTCATAGTCA 522
Db 48579 CAACGGTTTACTCGATTTCAGCAACCAATAC--ACTCGATTCTCTCATCACTTCTTCGTCC 48523

Qy 523 CTCGATCTCTCCTCACCATCTTTCATCATCTCCCTTATCTCGACCAACCGTCTCGCTCC 582
Db 48522 AAGCATCACTACTCTGACCT-TGCCTAATCACTCGACCT-----CACCATCTATCTTCGTCT 48468

Qy 583 ACCATTGCAATTTAAAAGCTCACTCGATTGTCAAAGAGAAAGAGTGAAGCTCAACAC 642
Db 48467 CTCATTTGCTCAACAAACAGTCATTTCTCCTCTCCAAGGAATACATCGACTCAACACA 48408

Qy 643 CGCCACTCGACCGGCTTTCCTCTACATTTCAACATTCGACGACCGGTGCTACCATCTCC 702
Db 48407 GCTTCTCTAGCTCATAACTCCACTCAACTTCACGCCAAGAGACCTCCGTTTCTCCACC 48348

Qy 703 AC-----ACCGCTCTTGTTCACCATACACTCGACCAACAACTCTCAAGTAAAA 752
Db 48347 ACCGATTCAACAACTCGAATCACCGCTCGTCTACCTCCACTCGGCTTCCGAAAACACTC 48288

Qy 753 AAAAAAGAAAAAAGTCAAAACCGACAGTTTCACTCAACACGGTTTACTCGACCGGTA 812
Db 48287 AACCCAGTTTCAAAAACCCGAAAAAANAANAANAANAANAANAANAANAANAANAANA 48228

Qy 813 CGCTGGTTTAGATTGTGTTTTGTTTTGTTTGTCTAATTAATAACATATTAAAGTTTATCTTGA 872
Db 48227 TACTCTTCTGTTGTTGTTGTTTTGTTTGTGTAACCTAACGTAATTAACGTTTATTATTGA 48169

Qy 873 GTTTCGTCTGTTTTTAGGTTTTCATGAGTAACCTACAGTGGAATAATCCTCTATGAGCCC 932
Db 48167 GTTACATTTGTTTTTTCAGGTTTTCATGAGTAACCTACAGTGGAATAATCCTCTATGAGCCC 48108

Qy 933 TCATTTAATGATGGGATGAAGCTAAGTCTCTGTTCTCACTAGACCGGAGTGAGAGCAACATGT 992
Db 48107 TGACTTACAATGTTGATGAGCTGAATCTCTGTTCCGCTAGACC--GAGAGAGCAACATGT 48051

Qy 993 TTACGAGAGCTATAGGGATGAATTTGAACGCTCTGACGCTCGACGTAATCAAGAGAGAC 1052
Db 48050 TTTCCAGAGCTTTCAGGGAAGAATTCGAACGCTCTGACGCGCCGCGTAATCAGAGAGAGC 47991

Qy 1053 TGAATCGCTAGAGGAAGAGGGCGATGTCGAGTAGATATGAGCTGATTGATCAGGATAT 1112
Db 47990 TGAATCGCCAGAGGAAGAGAGCGGATGTCGAAGATATGAGCTGATTGATCAGGAAAT 47931

Qy 1113 CAAAACGTGAGTATGAGCCAGAGTCATGGCGCAAGGAGAGCAAGACTACTGAACAAATCCGA 1172
Db 47930 CGAGACTGAGTATGAGCAAGAGTCATGGCGCAGGAGAGCAAGCTGCTGAACAAACCGGA 47871

Qy 1173 CGAGGTTACAGTGGAGGATATACAGATTCTTTGAGATGAATGACTTCTGGGAGACGAG 1232
Db 47870 CGAGGTTACAGTGGAGGATATATCAGGTTCTTTGAGGCTTAATGACTTCTAGGAGACGAG 47811

Qy 1233 GTATCCTGATATGAGCTTTAGCCAGTTGGGTTTACTCGAGGAGCTGACGATCTGTT 1292
Db 47810 GTATCCTGCTATGAGACTTTAGCCAGCTGGGGTTACTCGAGGAGCGTACAACTCTGTT 47751

Qy 1293 CGAAGAGTGTCTAT----- 1321
Db 47750 CGAAGAGTGTCTATGAGAGACTCATGTCTTACCTTATCTCTGCTTATAGGAGGAGAC 47691

Qy 1322 AATCGAGTTTCTTTCCACACTCGAAGTGGAATGATGAGGACTCACAGACTTTGAGCT 1381
Db 47690 AACCGAGTTTCTTTCCACCGCTGCAAGTGGAGATGATGAGGCACTCACAGACTTTGAGCT 47631

Qy 1382 GGATACCATGGGTTAGGCTTCTTTGACGTTCTTAGTGAGTGAAGAGCGGTAACAGATTGA 1441
```

|||||
Db 47630 GGATACCATGGGTTAGCGCTTCTGACGTTCTCAGTGAATGAACAGCGGTATGAGCTTTC 47571
|||||
QY 1442 GATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGGTT 1501
|||||
Db 47570 GATCAAGAAATTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGATT 47511
|||||
QY 1502 TGACAGGGAAGAGCTTAAGGATTTGTGGCTACTATTGCGGAACAATCACCCTAAACTC 1561
|||||
Db 47510 TGAAGAGAGAGGTTTAAGGATTTGTGGCTCACTATAGGGAACAATCACCCTAAACTC 47451
|||||
QY 1562 GACCGGTCCAAAGAGCAACCAAAATCCGAGTCTGTGATTCGCTACTTTACGCGCTCGGT 1621
|||||
Db 47450 GCGCGGTCTAAGAGCAACCAAAATCCGAGTCTGTGATTCGCTACTTTCAACCATCGGT 47391
|||||
QY 1622 TGCCAAATGTTTTTACTCCAGGAGTCTACAGGACCGGTCTAAACAGACATGAAGAT 1681
|||||
Db 47390 TGCCAAACGTCTTTACTCCAGGAGTCTACAGGACCGGTCTAAACAGACATGAAGAT 47331
|||||
QY 1682 GATAGATTCAGCGCTTATAGGATTTCCGCGCTTACAAAGGAAGAAATGTCCTGAGAGG 1741
|||||
Db 47330 GATAGATTTATGCGTTACAGGATTTCCGCGCTTACAAAGGAAGAAATGTCCTGAGAGG 47271
|||||
QY 1742 AGATCTTAAACACTCACACAGCAATGAGTAAAGAGGAGCACTATGCGTGGGTGGCTTGT 1861
|||||
Db 47270 AGATCTCAACAGCACACCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 47151
|||||
QY 1802 GTGGGCGCTGACAAACCGGCAAGAGAGGAGGAGCACTATGCGTGGGTGGCTTGT 1861
|||||
Db 47210 GTGGGCGCTGACAAACCGGCAAGAGAGGAGGAGCACTATGCGTGGGTGGCTTGT 47151
|||||
QY 1862 GAGCCCAATTCGAAAGTTTGTGGAGTTCGCTCAAGGAAGTGGGTTAGCACCGAAT 1921
|||||
Db 47150 GAGCCCAATTCGAAAGTTTGTGGAGTTCGCTCAAGGAAGTGGGTTAGCACCGAAT 47091
|||||
QY 1922 GATGGACTTGATCACTTCGCGCGATGTGAGTTCCTGAGTTTCACTGAGTTCGCGACTT 1981
|||||
Db 47090 GATGGACTTGATCACTTCGCGCGATGTGAGTTCCTGAGTTTCACTGAGTTCGCGACTT 47031
|||||
QY 1982 TCACCGCTACAGGTTTCAGGATTCATCGATTAGAAATCGCCAAACATTTCTTCCCTGCAT 2041
|||||
Db 47030 TCACCGCTACAGGTTTCAGGATTCATCGATTAGAAATCGCCAAACATTTCTTCCCTGCAT 46971
|||||
QY 2042 TTACGCTACTAGATTCCTGAGGCGAGGAAATGATCTCAAGCTTCGCTTGAAGATCT 2101
|||||
Db 46970 TGACGCTACTAGATTCCTGAGGCGAGGAAATGATCTCAAGCTTCGCTTGAAGATCT 46911
|||||
QY 2102 TTATTTTCGAGGCGAGTCCGCGCACTGAGGAGATTAGTCACCGAAGGAGCTACATAGA 2161
|||||
Db 46910 TTATTTTCGAGGCGAGTCCGCGCACTGAGGAGATTAGTCACCGAAGGAGCTACATAGA 46851
|||||
QY 2162 AGATTTGATGACATATGATATAGATGAGGCGAGTTTGACACGAGCATGTATCATTT 2221
|||||
Db 46850 AGATTTGATGACATATGATATAGATGAGGCGAGTTTGACACGAGCATGTATCATTT 46791
|||||
QY 2222 CAGTGAGCATATCTCCAGGAGGAAAGCAAGATTTGAGCGAAGCTCACAGGAACAA 2281
|||||
Db 46790 CAGTGAGCATATCTCCAGGAGGAAAGCAAGATTTGAGCGAAGCTCACAGGAACAA 46731
|||||
QY 2282 CAGCAAGCTGCAAGTGTGCAAGAAACAGGATAGTTTACTCGCAAGTGCCTCAGGCG 2341
|||||
Db 46730 CAGCAAGCTGCAAGTGTGCAAGAAACAGGATAGTTTACTCGCAAGTGCCTCAGGCG 46671
|||||
QY 2342 TATCAAGTTTCTGAAGCAAGATCAGTCTCTCTTCCCTTACCACTATTTCGCAATG 2401
|||||
Db 46670 TATCAAGTTTCTGAAGCAAGATCAGTCTCTCTTCCCTTACCACTATTTCGCAATG 46611
|||||
QY 2402 ACAGTCTCCTCAGGACATGCTTCGAGGAGATATGACGCGCC-----2444
|||||
Db 46610 ACATCTCCTCAGGACATGCTTCGAGGAGATATGACGCGCCCGGCGAGCGAG 46551
|||||
QY 2445 ----GAGCTCAGGAGCAGGATTCGATGCTCCCTGCGAGGATTCATCATTTGAGCC 2500
|||||

Db 46550 GCATGAGCTTAGAGCAGGAGATTCCGATGTCCCTGCGAGGATTTCATCATTCAGGCC 46491
|||||
QY 2501 TCCTGAAATCTAGGAGATAGGAAACGACACTCACTCGATCTAGCTAGCAGGAGGAGCAGC 2560
|||||
Db 46490 TCGTAAATCAGAGGAGCAGGAGGACCACTCGCTCGATCTAGCACTAGGAGGAGCAGC 46431
|||||
QY 2561 ACTTCTCAGTCTCGTGTAGTTTACCGCGCGGTGCTGGCCGCAATGAAGAAGAGAGGT 2620
|||||
Db 46430 ACTTCTCAGTCTCGTGTAGTTTACCGCGCGGTGCTGGCCGCAATGAAGAAGAGAGGT 46371
|||||
QY 2621 CGAGTATCTCAGAGCGGTGCTGGCCGCAATGAAGAAGAGAGGT 2680
|||||
Db 46370 CGAGTATCTCAGAGCGGTGCTGGCCGCAATGAAGAAGAGAGGT 46311
|||||
QY 2681 TGGAGCTGATACGGAACATGCGCGTTCGTATGCGCTTGGAGCAATCAGGAGGAGGT 2740
|||||
Db 46310 TGGAGCTGATACGGAACATGCGCGTTCGTATGCGCTTGGAGCAATCAGGAGGAGGT 46251
|||||
QY 2741 TGACTACCAACTTCGTTTCAATTCGAGTGAAGCGCTCACTTCCCATTTATTA 2800
|||||
Db 46250 TGAGGCAACTACGTTTCAATTCGAGTGAAGCGCTCACTTCCCATTTATTA 46191
|||||
QY 2801 TATCATCTCTGTGANTTGT-----CTTTATTTTGTTCAGTGAATTTGCTCTGA 2855
|||||
Db 46190 TACCATCTATGTGANTTGTTCCTTTCATTTTGTTCAGTGAATTTGCTCTGA 46131
|||||
QY 2856 GTACTCTCTTCCAGTTTATTCACAGTGGAGCTGTGATTTAAGTTTGGGAGGAGGT 2915
|||||
Db 46130 GTACTCTCTTCCAGTTTATTCACAGTGGAGCTGTGATTTAAGTTTGGGAGGAGGT 46071
|||||
QY 2916 CAGGAAGT--ATGTTGCAATGTATATATTTTAAAGTCTGATTCATCTCAAGGATACAAA 2973
|||||
Db 46070 CAGGAAGTGTGTTGCAATGTATATTTTAAAGTCTGATTCATCTCAAGGATACAAA 46011
|||||
QY 2974 AACCAAAAAAAT--AAAAATTCAGAAAATGATTTTAC--AAAAAAGAGTGTTCATGT 3031
|||||
Db 46010 AACCAAAAAAATTCAGAAAATGATTTTAC--AAAAAAGAGTGTTCATGT 45951
|||||
QY 3032 AGTTGATTAATCAATTTAGGATCAAGTCTAGAGTGTTCATTTAGGATTTGTCATGAT 3091
|||||
Db 45950 AGTTGATTAATCAATTTAGGATCAAGTCTAGAGTGTTCATTTAGGATTTGTCATGAT 45891
|||||
QY 3092 AGGGGATATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3122
|||||
Db 45890 AGGGGATATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 45860

RESULT 7

AB046433/c

LOCUS

DEFINITION

Arabidopsis thaliana DNA, chromosome 5 centromere region,

clone:Fl1C12.

ACCESSION

AB046433

VERSION

AB046433.1

KEYWORDS

centromere.

SOURCE

Arabidopsis thaliana

ORGANISM

Arabidopsis thaliana (thale cress)

Bukaryota; viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (sites)

Kumekawa, N., Hosouchi, T., Tsuruoka, H. and Kotani, H.

The size and sequence organization of the centromeric region of

Arabidopsis thaliana chromosome 5

DNA Res. 7 (6), 315-321 (2000)

JOURNAL

MEDLINE

PUBMED

AUTHORS

Kotani, H. and Kumekawa, N.

Direct Submission

Submitted (21-JUL-2000) Hirokazu Kotani, Kazusa DNA Research

Institute, Lab. Chromosome Research II; 1532-3 Yana, Kisarazu,

Chiba 292-0812, Japan (E-mail:kotani@kazusa.or.jp)

CDS	join(1601..1741,1851..1982,2226..2386,2480..3428)
	/gene="Tl17A2.2"
	/note="contains similarity to Pfam family PF00155 - Amino transferases class-I; score=584.1, E=8.8e-172, N=1; strong similarity to 1-aminocyclopropane-1-carboxylic acid synthases"
	/codon_start=1
	/evidence=not experimental
	/protein_id="AAD48074.1"
	/db_xref="GI:5724770"
	/translation="MLSSKVGVDSHGQDSYFLGWQYKPNFHSFNTSGIVQMGLA ENQLSFDLIEKWLKEEPEVLGLKKNDESVERQALFQDYHGLPAFADMAKFMGKIRE NKVFDTNKMVLTAGTSANETLMFLCLANFGDAFLIPAPYPGFDRDLKWTGVEIIVP IHCVSSNGYKITEDALEDAYERALKHNLNVKGVLTITNPSPLGSTRRELDLLLTFT STKIKHMSDEIYSGTVFDSPEPTSVILEVAKDNMGDKIHHVYSUKDGLGEGFRV LVIYNNKVVSAATKMSFGLISSQTHLLANLLSDERTTYLLEENKRLERKDR GLVSLKEAGISCLKSNAGLFCWDLRHLLKSTFEAHSMTKIVCEVGLNISPSSC HCDPEGFRVCFANMSQITMEVANDRVKGFVDNNGGKQKRTMTDTRRSRLINKVSK LSSVTCESER"
repeat_region	complement(6823..6962)
	/note="Athila retroelement"
repeat_region	complement(7214..7279)
	/note="Athila retroelement"
repeat_region	complement(7237..7358)
	/note="106B"
repeat_region	complement(7299..7330)
	/note="Athila retroelement"
repeat_region	complement(7361..7480)
	/note="106B"
repeat_region	complement(7470..7548)
	/note="Athila retroelement"
repeat_region	complement(7473..7583)
	/note="106B"
repeat_region	complement(7581..7631)
	/note="106B"
repeat_region	complement(7585..7620)
	/note="Athila retroelement"
repeat_region	complement(7631..7654)
	/note="106B"
repeat_region	complement(7692..7718)
	/note="Athila retroelement"
repeat_region	complement(8439..8722)
	/note="Athila retroelement"
repeat_region	complement(8792..8828)
	/note="Athila retroelement"
repeat_region	complement(8802..8828)
	/note="H12 Athila 180 bp_repeat"
repeat_region	complement(8826..8881)
	/note="Athila retroelement"
repeat_region	complement(8826..8881)
	/note="H12 Athila 180_bp_repeat"
repeat_region	complement(8882..8913)
	/note="Athila retroelement"
repeat_region	complement(8882..8913)
	/note="H12 Athila 180_bp_repeat"
repeat_region	complement(8919..9023)
	/note="H12 Athila 180_bp_repeat"
repeat_region	complement(8934..9023)
	/note="Athila retroelement"
repeat_region	complement(9380..9446)
	/note="Athila retroelement"
repeat_region	complement(9444..9600)
	/note="H12 Athila 180_bp_repeat"
repeat_region	complement(9444..9584)
	/note="Athila retroelement"
repeat_region	complement(9603..9661)
	/note="Athila retroelement"
repeat_region	complement(9603..9661)
	/note="H12 Athila 180_bp_repeat"
repeat_region	complement(9667..9751)
	/note="Athila retroelement"
repeat_region	complement(9667..9750)
	/note="H12 Athila 180_bp_repeat"

LOCUS	Tl17A2	81902 bp	DNA	linear	PLN 10-AUG-1999
DEFINITION	Arabidopsis thaliana BAC Tl17A2.				
ACCESSION	AF160183				
VERSION	AF160183.1 GI:5106767				
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Arabidopsis thaliana				
REFERENCE	AUTHORS				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS	TITLE				
JOURNAL	REFERENCE				
AUTHORS					

	Query Match	62.0%	Score 1937.2;	DB 8;	Length 81902;
	Best Local Similarity	84.8%;	Pred. No. 0;		
	Matches 2289;	Conservative	0;	Mismatches 343;	Indels 66;
				Gaps	8;
Qy	483	CAACCAAACTCGACCTCGTCTCTCTTTGGCACTCATAGTCACTCGATCTCTCTCTCAACATC	542		
Db	12553	CAACTTCACCTCGAACGGTCTCTCTCTGGCTTTTCACAGTCACCTCGATCTGGCCAAAGCCA-C	12495		
Qy	543	TTCAATCATCTCCCTTTACTCGACCAACCGTGGTCTCGCTCCACATTTGCCATTTAAAGCT	602		
Db	12494	ACCATCTCTCCATCTCTTTTGCACTCGATCGGGTCTACACACCATCAACAACACTCTCGAC	12435		
Qy	603	CACCTCGATTCTCAAGAGAGAGAGTGAAGTCAACACACCGGCACTCGACCGGTTTCC	662		
Db	12434	CTCCTGCTCTTTCTCAAGGAGACAGCTCAACTCGCGGTTTGATTCATCTCGGCCACAGC	12375		
Qy	663	CTCTACACATTCAACACTCGACCAACGGTGTACCATCTCCACACCGGCTCTTGTTTCACCA	722		

QY	1769	GCCTCTCTGATCCATCTGTGTGGTACATGAAGTGGCGCTGACAAACGCGCAAGAGAA	1828
Db	11241	GCCTCTATATCCATCTGTGTGGATACAGGAAGTGGCGCTGACGAACGGAAGAGAA	11182
QY	1829	GGTAAAGGAGCACTATGCGTGGGTGGTGTGACGCCAATTCCTGAAAGTTTCTGGAGT	1888
Db	11181	GGTGAGAGGAGCACTATGCGTGGGTGGTGTGACGCCAATTCCTGAAAGTTTCTGGAGT	11122
QY	1889	TCCGCTCAAGGAAGTAGGGTTAGCACCGAGAAATGATGACCTTGGATCATCTTGGCCGATG	1948
Db	11121	ATCCGCTCAAGGAACACAGGGTTAGCACCGAGGAATGATGACCTTGGATCATCTTGGCCGATG	11062
QY	1949	TGAGTTCTCTGAGTTTGACATGGTTGGGACCTTTCACCGCTACAGCTTCGAGCAATTCATC	2008
Db	11061	TGAGTTCTCTGAGTTTGACATGGCTGGCGACCTTTCACCGCTATAGTTTCGAGCAATTCATC	11002
QY	2009	GATTAGATCCCAACATCTCTTTCCCTCGCATTTACGCTACTAGGATTTCTCGAGGGCAG	2068
Db	11001	GATTAGAAATGGCAACATCTCTCTCCCTGCATTGCTACTAGGATTTCTCGAGGGCAG	10942
QY	2069	GAACATTCAGTTCAAGCTCGGCTTGAAGATCTTTATTTTCAGGGCAGTCCGCCAATCGA	2128
Db	10941	GAACATTCAGTTCAAGCTCGCACTTGAAGATCTTTATTTTCAGGGCAGTCCGCCAATCGA	10882
QY	2129	GGAGATTAGTCACACCGAAGGAGCTACAATGAAGATGTTGATGACATATGATATAGA	2188
Db	10881	TGAGATCATGTACACCGAAGGAGCTACAACAGAGATGTTGATGACACGATATAGA	10822
QY	2189	TGAGCGGAGTTTGACACGAGCATGTATCATTTTCAGTGAACATATACCTCCAGCGAGAA	2248
Db	10821	TGAGCGGAGTTTGACACGAGTATGTATCATTTTAGTGACATATACCTCCAGCGAGAA	10762
QY	2249	AAGCAAGATTTGACGAGCTCACAGGAACAAACAGCAAGCTGCAGAGTGTGCAGAA	2308
Db	10761	AAGCAAGATTTGAGCGAAGCTCACAGGAACAAACAAAGCTGCAGAGTGTGCAGAA	10702
QY	2309	ACAGATATAGTTACTCGCCCAAGTCTCAGGGCTATCAAGTTTCTGAAGCAAGATATAG	2368
Db	10701	ACAGGATAAGTTTACTCGCTAAGTCTCAGGGCTATCAGGTTTCTGAAGCAAGATATAG	10642
QY	2369	CTGCTCTCTTCCACTACAACTATTCGCAATTCAGAGCTCCCTCAGGACATGCTTCGAG	2428
Db	10641	CTGCTCTCTTCCACTACAGCTATTCGTTAGGACAGCTCCCTCCGACATGCTTCGAG	10582
QY	2429	GAGATATGACGGCCC-----GAGCCTAGAGCAGAGATTTCT	2467
Db	10581	GAGATATGACGGCCCGAGCCAAAGTCAGCGCAGGCTGAGCCTAGAGCAGGAGATTC	10522
QY	2468	GCATGTCCCTGGAGGCAATTCATTCGAGCCTCGTGAATCTAGGAAGAAATAGGAGAAC	2527
Db	10521	GCATGTCCATCGAGAGCAATTCATTCGAGCCTCGTGAATCTGGGAGGAGGAGGAC	10462
QY	2528	GACACTCACTCGATCTAGCAGCAGGAGCAGACGACTTCTCGAGTCTCGTAGTTACCGGA	2587
Db	10461	CACACTCACTCGATCTAGCAGCAGGAGCAGACGACTTCTCGAGTCTCGTAGTTACCGGA	10402
QY	2588	CCGCGGTGCTGGCGCAATAGAGAGAGAGTTCGATATCTTCAGAGCGGTCTGGCCG	2647
Db	10401	CCGCGGTGCTGGCGCAGAGAGAGAGGTTAAGTATCTTCAGAGCGGTCTGGCCG	10342
QY	2648	CCACAGAGCTGATGAGATCCAGTACCCACATGCTGGAGCTGATACGGAACATGCGGTTTC	2707
Db	10341	CCACAGAGCTGATGATGTCGAGTACCCACCTGCTGGAGCTGATACAGAACAGGTGTTTC	10282
QY	2708	GTCTATGGCTGGGAGCAATCACAGGAGCAGCAATGATACCAATCTGCTTCATTATCGA	2767
Db	10281	GTCCATGCTGGGAGCAATCACAGGAGCAGCAATGATGAGCAATACGTTTCATTCTCGA	10222
QY	2768	CTGAGGTAAGCGCTCACTTACCAATATATATATATATATATATATATATATATATATAT	2826
Db	10221	CTGAGGTAAGCGCTCACTTACCANTGTATATATATATATATATATATATATATATATAT	10162
QY	2827	TTTTTTTTCAGTTCGATTCGATTTGCTCCTGAGTACTCTCTCCAACTTTATTTCACACAGTGG	2886
Db	10161	TTTTTTTTCGATTCGATTCGATTTGCTCCTGAGTACTCTCTCCAACTTTATTTCGACACAGTGG	10102
QY	2887	ACTGTGTGATTAAAGTTTGGGGAGGGCTCAGGAAGTATG--TTGCATTGTATATATTTT	2944
Db	10101	ATTGTGTGATTAAAGTTTGGGGAGGGCTCAGGAAGTGTGTATTGCAATTTGTATATATTTT	10042
QY	2945	TAACTCTGCATTCATCTTAAGCATAGAAAAACCAAAAAAATTTAAAAAATTTTCAGAAAT	3004
Db	10041	TGAGTCTGCATTCATCTTAAGCATAGAAAAACCAAAAAAATTTAAAAAATTTTCAGAAAT	9982
QY	3005	GATTTCACAAAAAAGAGTGTTCATGTAGTTTGCATTACATTAGCATCAAGTCTAGAGTG	3064
Db	9981	GATTTCACAAAAAAGAGTGTTCATGTAGTTTGCATTACATTAGCATCAAGTCTAGAGTG	9922
QY	3065	TTTCATTTAGGATTTGTCATATGATAGGGGATTAATGATGAGATAGCTTGTGAAGCA	3122
Db	9921	TTTCATTTAGGATTTGTCATATGATAGGGGATGATGATGAGATAGCTTGTGAAGCA	9864
RESULT 9			
ATCHRIV21/c 195837 bp DNA linear PLN 16-MAR-2000			
LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 21.			
DEFINITION AL161509			
ACCESSION AL161509.2 GI:7267442			
VERSION			
KEYWORDS			
SOURCE Arabidopsis thaliana (thale cress)			
ORGANISM Arabidopsis thaliana			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
1 (bases 42621 to 43431)			
REFERENCE Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and			
Mayer,K.F.X.			
JOURNAL Unpublished			
REFERENCE 2 (bases 43282 to 43283)			
Lamar,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and			
Mayer,K.F.X.			
JOURNAL Unpublished			
REFERENCE 3 (bases 109648 to 109649)			
Spiegel,L.A., Huang,E.N., Nascimento,L.U., de la Bastide,M.,			
Vil,D.M., Preston,K.R., Matero,A., Shah,R., O'Shaughnessy,A.,			
Rodriguez,M., Shekher,M., Schutz,K., See,L.H., Swaby,I.,			
Habermann,K., Dedhia,N.N., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.			
JOURNAL Unpublished			
REFERENCE 4 (bases 1 to 195837)			
EU Arabidopsis sequencing,project.			
JOURNAL Direct Submission			
REFERENCE Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer			
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:			
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,project			
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge			
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,			
E-mail: michael.bevan@bbsrc.ac.uk			
COMMENT Information on performance of analysis and a more detailed			
annotation of this entry and other sequences of chromosomes 3, 4			
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/			
this fragment has an overlap with ATCHRIV20 at the 5' end and an			
overlap with ATCHRIV22 at the 3' end.			
FEATURES			
source			
1. .195837			
/organism="Arabidopsis thaliana"			
/mol_type="genomic DNA"			
/variety="Columbia"			
/db_xref="taxon:3702"			
/chromosome="4"			
12352..12656			
/note="function=repeat_mil67, GenBank accession number X94367"			
12651..12915			
/note="function=repeat_mil67, GenBank accession number X94367"			
repeat_region			
repeat_region			

[illegible]

```
|||||
53803 GCATGTCATCGGAGACATTCATTCATTCGAGCCTCGTAAATCTGGAGGAAGAGGAGGAC 53744
|||||
2528 GACACTACTCGATCTAGCAGCAGGAGCAGACGACTTCTGCGAGTCTCGTAGTTTACGCGA 2587
|||||
53743 CACACTACTCGATCTAGCAGCAGGAGCAGACGACTTCTGCGAGTCTCGTAGTTTACGCGA 53684
|||||
2588 CCGCGGTGCTGGCGGCAATAGAGAAGAGAGGTGAGTATCCTCAGACGGTCTGGCGG 2647
|||||
53683 CCGCGGTGCTGGCGGCAATAGAGAAGAGAGGTGAGTATCCTCAGACGGTCTGGCGG 53624
|||||
2648 CCACAGAGCTCATGAGATCGAGTACCCACATGCTGGAGCTATACGGAACATGCGGTTTC 2707
|||||
53623 CCACAGAGCTCATGATGTCGAGTACCCACCTGCTGGAGCTATACGGAACATGCGGTTTC 53564
|||||
2708 GTCTATGGCTGGAGCAATCACAGGAGCAGCATTGACTACCACTCTGTTCAATATTGGA 2767
|||||
53563 GTCCATGGCTGGAGCAATCACAGGAGCAGCATTGACTACCACTCTGTTCAATATTGGA 53504
|||||
2768 CTGAGGTAAGCGCTCCTACCTCACCATTATATATATATATATATATATATATATATAT 2826
|||||
53503 CTGAGGTAAGCGCTCCTACCTCACCATTATATATATATATATATATATATATATATAT 53444
|||||
2827 TTTTGTTTCAGTGATTCGATTGCTCCTGAGTACCTCTCCAGTCTCTTCCAGTCTTATTC 2886
|||||
53443 TTTTGTTTCAGTGATTCGATTGCTCCTGAGTACCTCTCTTCCAGTCTTATTC 53384
|||||
2887 ACTGTGTGATTAAGTTTGGGGAGGCTCAGGAAGTATG--TTGCAATTGATATATTTT 2944
|||||
53383 ATTGTGTAATTAAGTTTGGGGAGGCTCAGGAAGTATGTTGATTCATTTGATATATTTT 53324
|||||
2945 TAAGTCTGCATTCATCTAAGGCATAGAAAAACAAAAAAATTTTCAAGAAAT 3004
|||||
53323 TGAGTCTGCATTCATCTAAGGCATAGAAAAACAAAAAAATTTTCAAGAAAT 53264
|||||
3005 GATTTCACAAAAAAGAGTGTTCATGTAGTTGCAATACATTTAGGATCAAGTCTAGAGTG 3064
|||||
53263 GATTTCACAAAAAAGAGTGTTCATGTAGTTGCAATACATTTAGGATCAAGTCTAGAGTG 53204
|||||
3065 TTTCAATTAGGATTTGTCATATGCATAGGAGGATATGATGAGATAGCCTTTGAAGCA 3122
|||||
53203 TTTCAATTAGGATTTGTCATATGCATAGGAGGATGATGATGATAGCCTTTGAAGCA 53146

RESULT 10
AB073166
LOCUS
DEFINITION
Arabidopsis thaliana DNA, chromosome 4 centromere region, BAC
clone:F15N16.
ACCESSION
AB073166
VERSION
AB073166.1 GI:18149207
KEYWORDS
Arabidopsis thaliana (thale cress)
SOURCE
Arabidopsis thaliana
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 Kumeoka,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.
The size and sequence organization of the centromeric region of
Arabidopsis thaliana chromosome 4
DNA Res. 8 (6), 285-290 (2001)
JOURNAL
MEDLINE
21842138
PubMed
11853315
REFERENCE
2 (bases 1 to 123460)
Kotani,H. and Kumeoka,N.
Direct Submission
Submitted (16-OCT-2001) Hirokazu Kotani, Kazusa DNA Research
Institute, Lab. Chromosome Research II; 1532-3 Yana, Kisarazu,
Chiba 292-0812, Japan (E-mail:kotani@kazusa.or.jp,
Tel:81-438-52-3920, Fax:81-438-52-3921)
LOCATION/Qualifiers
1. .123460
FEATURES
Source
```

```

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
/map="centromere"
/clone="F15N16"
1. .1000
/feature="overlap to BAC F13J5, please refer for analysis and
annotation"
1000. .5529
/feature="similarity to fragments of Athila retrotransposons"
5530. .7082
repeat_region
7083. .20826
/feature="similarity to 180 bp centromeric repeat"
20827. .24935
/feature="similarity to 160 bp repeat"
24936. .25193
LTR
25194. .27048
27049. .34345
misc_feature
34346. .35940
/feature="similarity to fragments of Athila retrotransposons"
35941. .41422
/feature="similarity to fragments of Athila retrotransposons"
41423. .43041
LTR
43042. .44457
LTR
44458. .45957
misc_feature
45958. .53796
/feature="similarity to fragments of Athila retrotransposons"
53797. .55873
LTR
55874. .58433
repeat_region
58434. .58805
58807. .62369
misc_feature
/feature="similarity to fragments of Athila retrotransposons"
62370. .62737
LTR
62738. .78607
repeat_region
/feature="similarity to 160 bp repeat"
complement(79991..81205)
/codon_start=1
/evidence=not_experimental
/product="extensin-like protein"
/protein_id="BAB83614.1"
/db_xref="GI:18149208"
/translation="MNHLLALFLVGTIALPDRGTQQLTNDGLTGARREILSFQDR
LAVYPIORFKSLTLDIPNVTKTWIGSDICSYRGHCDNPPHNTAVTASIDFNG
FQSPASISGFIQFADLALFHVNSNFGTVPSTKIVNLYLYELDISNNRFTGPPT
AVGMSGLTFIDIRFNSFGSIPQILQNLVFLINDNGFTASLEIPDQGTTHILF
LTIANKNFGLPLRSILRSMSTLTVLFLNNDFTGCIPIHEIGFTCASVIDIGNKLT
GPLELSLMCLEVQDLNFAAGLLFGAVPEAVCMLLDLNVLNLSNLYTHVCPMCRG
LLEKGVLDVGNNCIPFPQGRSMQCAEEFVFKPKYKCPHMHFHSFSCRYSHISSS
SSAFPMWVAPSP"
complement(83214..83666)
/codon_start=1
/evidence=not_experimental
/product="Ap2 domain containing protein"
/protein_id="BAB83615.1"
/db_xref="GI:18149209"
/translation="MVIOYRKQEPFMVKEGMVMTKPKRNLISSNEKYKGIKRMK
GKWAIRREPKNRSIRLWGYKTAAVAARAYDTAVFYLRGSPARLNFPEVFKDNGG
EGLGDMSPFTLIRKKAEEVGARVDAELRENWENLDNMKLPAYGL"
93163. .95514
misc_feature
/feature="similarity to fragments of En/Spm transposon"
ORIGIN
Query Match 61.7%; Score 1925.4; DB 8; Length 123460;
Best Local Similarity 89.2%; Pred. No. 0;
Matches 2203; Conservative 0; Mismatches 201; Indels 67; Gaps 9;
QY 716 TTCACATACATCGACCACTCTCAAGTAAAAAAGAAAAAAGTCAA 775
DB 50062 TTCCATAATCACTCGATCACTACCCATAAAGAAAAAATTTTCA 50121
```


Db 52279 TTCGACAGTGGACTGTGTGATTAAAGTTTGGGGAGGGCTCAGGAAGTGTGTGTGCAT 52338

QY 2933 TGTATATATTTTAACTGTGCATTCATCTAAGGCATAGAAAAACCAAAAAAATTTAAAA 2992

Db 52339 TGTATATATTTTGAAGTGTGCATTCATCTAAGGCATAGAAAAACCAAAAAAATTTGAAA 52398

QY 2993 ATTTCAAGAAATGATTTTAC-AAAAAAGAGTGTTCATGTAGTTGCATTCATTTAGAT 3051

Db 52399 ATTTCAAAAAAGATTTTACAAAAAAGAGTGTTCATGTAGTTGTATTACATTTAGAT 52458

QY 3052 CAAGCTAGAGTGTTCATTTTAGCATTTGCATATGCATAGGGGATAAATCATCAGATAG 3111

Db 52459 CGATTTACAGTGTTTATTATTAGGATTTGTCATATACATAGGGGATAAATCATCAGATAG 52518

QY 3112 CCTTGTAAGCA 3122

Db 52519 CCTTGTAAGCA 52529

RESULT 11

AC084241/c

LOCUS

DEFINITION

AC084241

VERSION

AC084241.1 GI:10864177

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids; 1 (bases 1 to 14879)

Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J., Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.

Arabidopsis thaliana chromosome 1 BAC F16M11 genomic sequence

Unpublished

2 (bases 1 to 14879)

Town,C.D. and Kaul,S.

Direct Submission

Submitted (18-OCT-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org

3 (bases 1 to 14879)

Town,C.D. and Kaul,S.

Direct Submission

Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org

Address all correspondence to:at@tigr.org

COMMENT

BAC clone F16M11 is from Arabidopsis thaliana chromosome 1

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkEM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pertea,

<http://www.tigr.org/softlab/glimmerm.htm>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

source

Location/Qualifiers

1. .14879

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="I"

/map="?"

/clone="F16M11"

complement(1..1054)

/rpt_family="Rf:gi|7330786|gb|AF204874.1|AF204874

Arabidopsis thaliana conserved centromere sequence CCS1"

1

/note="(TR) 60155 nt beyond this point were not included in the submitted sequence due to an overlap with another BAC (F19C17)"

65..5003

/rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila retroelement.#Rf:ATRO033|X81801 Athila retroelement"

complement(149..1157)

/rpt_family="Rf:gi|7330787|gb|AF204875.1|AF204875

Arabidopsis thaliana conserved centromere sequence CCS2"

complement(149..417)

/rpt_family="Rf:centromeric repeat from T7M24 72424 to 72761"

complement(1271..1403)

/rpt_family="Rf:gi|7330786|gb|AF204874.1|AF204874

Arabidopsis thaliana conserved centromere sequence CCS1"

complement(1465..1677)

/rpt_family="Rf:gi|7330787|gb|AF204875.1|AF204875

Arabidopsis thaliana conserved centromere sequence CCS2"

1805..2183

/rpt_family="Rf:ATGAGPOL1_I|ATGAGPOL1_I a consensus. similar to del-like retrotransposon from T32N15, GB|AC002534."

1805..2180

/rpt_family="Rf:ATGP2I|ATGP2I Internal region of ATGP2 gypsy-like LTR retrotransposon."

1805..2180

/rpt_family="Rf:TY3G|TY3G Ty3/gypsy LTR retrotransposon (a consensus)."

1807..3349

/gene="F16M11.1"

/note="pseudogene, gypsy/Ty3-element polyprotein (fragment)"

/pseudo

complement(1832..2170)

/rpt_family="Rf:ATRO059|AC002534 T32N15 del-like retrotransposon"

1837..2000

/rpt_family="Rf:TEKAYI|TEKAYI Internal part of gypsy/Ty3-type retrotransposon Tekay."

2819..3171

/rpt_family="Rf:ATGP2I|ATGP2I Internal region of ATGP2 gypsy-like LTR retrotransposon."

2828..3261

/rpt_family="Rf:ATGAGPOL1_I|ATGAGPOL1_I a consensus. similar to del-like retrotransposon from T32N15, GB|AC002534."

complement(2843..3267)

/rpt_family="Rf:ATRO059|AC002534 T32N15 del-like retrotransposon"

2843..3261

/rpt_family="Rf:TY3G|TY3G Ty3/gypsy LTR retrotransposon (a consensus)."

3349..5004

/rpt_family="Rf:ATHILA|ATHILA A.thaliana DNA for Athila retroelement.#Rf:ATRO033|X81801 Athila retroelement"

3349..5003

/rpt_family="Rf:Athila LTR for Athila retroelement"

3349..3780

/rpt_family="Rf:gi|895701|emb|X89195.1|ATSATDNA1

A.thaliana DNA of a 180bp satellite

repeat_region

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

gene

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

Db	11709	AAGAAATGCTTGGAGGCGATCTCAACAACGACCACCACTATGCTCTGTTGATCCAC	11650
Qy	1785	CTGTCTGGGTACATGAAGTGGCGCTGACAAACGCAAGAAGAGTAAGAGG-AGCACT	1843
Db	11649	CTGTGTGGATACAGGAAGTGGCGGTTGATAAACGGCAAGAAAGTGGAGGCGAGCACT	11590
Qy	1844	ATGCGTGGGTGGCGTCTGTGACGCCAAATTCGTAAAGATTGTGGAGTTCCGCTCAAGAAAGT	1903
Db	11589	ATGCGTGGGTGGCGCTGTGATGTCCAATTTTGGAAAGCTTGTGGAGTACCGCTCAAGAGCC	11530
Qy	1904	AGGGTTAGCAACCCAGAGATGATGACTTGGATCACTTGCGCCGATGTGAGTTCCTGTAGTT	1963
Db	11529	AGGGTTAGCAACCCAGAGATGATGACTTGGATCACTTGCCTGATGTGAGTTCCTGAGTT	11470
Qy	1364	TGACATGGTTGGCGCACTTTCAACCGCTACAGGTTCCGAGCAATTCATCGATTAGAAATCGCCAA	2023
Db	11469	TGACATGGTTGGCGCACTTTCAACCGCTACAGATTCGAGCAATTCATCGATTTCGAATCGCCAA	11410
Qy	2024	CATTCTTTTCCCTGCAATTTACGCTACTAGGATTCGAGGGCGAGCAATTCGATTCGAA	2083
Db	11409	CATTCTTTCTCCCTGCAATTTGACGCTACTAGGATTCGAGCAAAAGGCAATTCGATTCGAA	11350
Qy	2084	GCCTGGCCTTGAAAGATCTTTATTTTCGAGGGCGAGTCGCCCACTGAGGAGATTAFTCAAC	2143
Db	11349	GCCTGGCCTTGAAAGATCTTTATTTTCGAGGGCGAGTCGCCCACTGAGGAGATTAFTCAAC	11290
Qy	2144	CGAAGGAGCTACAAATAGAAGATGTTGATCAGACATATGATATAGATGAGGCGGAGTTTGA	2203
Db	11289	CGAAGGACCTACAAAGAGAGATGTTGATCGAGACAGATGATGATGAGGCGGTTTGA	11230
Qy	2204	CACGAGATGTATCATTTTCAGTGAGCATATACCTCCAGCGAGAAAGCAAGAGTTTGAG	2263
Db	11229	CACGAGCATGTATCATTTTGTAGTGAGCATATACCTCCAGCGAGAAAGCAAGAGTTTGAG	11170
Qy	2264	CGAAGCTCACAGAAACAAACAGCAAGCTGCAGAAAGTGGTGCAGAAACAGGATTAAGTACT	2323
Db	11169	CGAATCTCAGAGAACAATAGCAAGCTGCAGAAAGTGGTGCAGAAACAGGATAAGTACT	11110
Qy	2324	CGCCAAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAAGATCAGCTGCTCCTCTCCAC	2383
Db	11109	CGCCAAATGCTCTCAGGGCTATCCAGTTTCTGAAGACAAAGATCAGCTGCTCCTCTCCAC	11050
Qy	2384	TACAACTATTTCGCAATGACAGCTCCCTCAGGACATGCTTTTCGAGGAGATATGACGGCC	2443
Db	11049	TACAGCTATTTCGCGAGGACAACTCCCTCGGACATGCTCCCTCGAGGAGATATGACGGCC	10990
Qy	2444	CGAGCC-----TAGAGACGAGAAGATTCGTGATGTCCCTGCGAG	2482
Db	10989	TGAGCCAAAGTCGGCGAGGCTTGAGCATAGAGAGCAGGAGATTTTCGATGTCCCTGCGAG	10930
Qy	2483	GCATTTCATCATTTTCGAGCCTCGTGAATCTAGGAAGATAGGAGAACGACACTCACTCCATC	2542
Db	10929	GCATTTCATTTCGAGCCTCGTGAATCAGGAGAGAGTGGAGAGCACTCACTCCATC	10870
Qy	2543	TAGCAGCAGGAGCAGACACTTCTCGAGTCTCGTAGTTTACGCGACCGCGTGTGCGCG	2602
Db	10869	TAGCAGCAGGAGCAGACACTTCTCGAGTCTGTAGTTTACGCGACCGCGTGTGCGCG	10810
Qy	2603	CAATAGAGAAGAGAGGTTCGATATCTCAGAGCGGTGCTGGCGCCACAGCTGATGA	2662
Db	10809	CAGAAGAAGAAGAGGTTCGATATCTCAGAGCGGTGCTGGCGCCACAGCTGATGA	10750
Qy	2663	GATCGAGTACCCACATGCTGGAGCTGATACGGAACATGGCGGTTTCGCTATGCTTGGGA	2722
Db	10749	SGTCGAGTACCCACCTGCTGGAGCTGATACAGAACAGGCGAGTTTCGCTATGCGCTGGGA	10690
Qy	2723	GCAATCAGGCGACCATTCGACTACCAACTTCGTTTCATTTTCGACTGAGTAAGCGCT	2782
Db	10689	GCAATCAGGCGACCATTCGACTACCAACTTCGTTTCATTTTCGACTGAGTAAGCGCT	10630
Qy	2783	CACTTCCACATTATATATCATCTCTGTGATTTGTTTCTTTATTTTGTTCAGTGAT	2842
Db	10629	CACTTCCACATTATATATCATCTCTGTGATTTGTTTCTTTATTTTGTTCAGTGAT	10570

RESULT	12	ACCESSION	REFERENCE	TITLE	REFERENCE	COMMENT
AC073433	VERSION	AUTHORS	JOURNAL	AUTHORS	AUTHORS	
LOCUS	KEYWORDS	SOURCE	REFERENCE	TITLE	TITLE	
DEFINITION	ORGANISMS			JOURNAL	JOURNAL	

RESULT 12	AC073433	117055 bp	DNA	linear	PLN 19-JAN-2001
LOCUS	AC073433				
DEFINITION	Arabidopsis thaliana chromosome 1 BAC F19C17 genomic sequence, complete sequence.				
ACCESSION	AC073433				
VERSION	AC073433.9	GI:12321079			
KEYWORDS	HTG.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi				
REFERENCE	1. (bases 1 to 117055)				
AUTHORS	Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J., Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.				
TITLE	Arabidopsis thaliana chromosome 1 BAC F19C17 genomic sequence				
JOURNAL	Unpublished				
REFERENCE	2. (bases 1 to 117055)				
AUTHORS	Town, C.B. and Kaul, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUN-2000) The Institute for Genomic Research, 9712				
REFERENCE	3. (bases 1 to 117055)				
AUTHORS	Town, C.D. and Kaul, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-JAN-2001) The Institute for Genomic Research, 9712				
COMMENT	Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org On Jan 19, 2001 this sequence version replaced gi:12280781. Address all correspondence to: atigr.org				

BAC clone F19C17 is from Arabidopsis thaliana chromosome 1. The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/Genemark/>), GlimmerA (a variant of GlimmerM, see Mihaela Perlea, <http://www.tigr.org/sortlab/glimmerm.htm#glimmerm.html>, and GeneSplicer (Mihaela Perlea and Steven Salzberg, contact mpereate@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as

hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

source	1. 117055	/organism="Arabidopsis thaliana"	/rpt_family="Rf:ARNOLD2 ARNOLD2 Autonomous DNA-transposon
		/mol_type="genomic DNA"	ARNOLD2."
		/cultivar="Columbia"	complement (8636. .8980)
		/db_xref="taxon:3702"	/rpt_family="Rf:ARNOLD3 ARNOLD3 Autonomous DNA-transposon
		/chromosome="1"	ARNOLD3."
		/clone="F19C17"	complement (8636. .8980)
repeat_region	103. .397	/rpt_family="Rf:centromeric repeat from T7M24 69183 to 70294"	ARNOLD4 Autonomous DNA-transposon
		/rpt_family="Rf:centromeric repeat from T7M24 69183 to 70294"	ARNOLD4."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement.#Rf:ATRO033 X81801 Athila retroelement"	complement (8636. .8878)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement.#Rf:ATRO033 X81801 Athila retroelement"	ARNOLD2 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:gi 895704 emb X89198.1 ATSATDNA4 A.thaliana DNA of a 180bp satellite junction#Rf:ATRO037 X89198 clone H33 of Athila 180-bp repeat"	complement (8636. .8878)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement.#Rf:ATRO033 X81801 Athila retroelement"	ARNOLD4 Autonomous DNA-transposon
repeat_region	305. .397	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement.#Rf:ATRO033 X81801 Athila retroelement"	ARNOLD1 Autonomous DNA-transposon
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD1."
repeat_region	305. .397	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8636. .8875)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8640. .8888)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD2 Autonomous DNA-transposon
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD2."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8640. .8883)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD1 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD1."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8640. .8812)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3 Autonomous DNA-transposon
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8641. .9231)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8641. .8809)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8660. .8807)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8667. .24255)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8672. .9425)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4 Autonomous DNA-transposon
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8681. .8795)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD2 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD2."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8683. .8795)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD1 Autonomous DNA-transposon
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD1."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8684. .8805)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8685. .9190)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8685. .9148)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD4."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8685. .8883)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3 Autonomous DNA-transposon
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD3."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8686. .8771)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8686. .8765)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	transposon ARNOLDY1 - a consensus sequence."
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8687. .8875)
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD1 Autonomous DNA-transposon
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD1."
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	complement (8688. .8809)
		/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD2 Autonomous DNA-transposon
repeat_region	305. .399	/rpt_family="Rf:ATHILA ATHILA A.thaliana DNA for Athila retroelement"	ARNOLD2."

ARNOLD2. complement (8689...8849) /rpt family="RF:ARNOLD1 ARNOLD1 Autonomous DNA-transposon ARNOLD1." repeat_region complement (8689...8849)		Query Match Best Local Similarity 61.2%; Score 1911.6; DB 8; Length 117055; Matches 2147; Conservative 0; Mismatches 174; Indels 61; Gaps 7;
QY	796	GGTTTACTCGACCGGTACGCTGGTTTAGATTCGTGTTTTGAGTCTGTTTTGGTTTGCTATTACTACATCA 855
Db	112558	GGTATACTCGACCGATTTACTCTGTGGTTTGTTTTGGTTTAGTGGGATAACGTA 112617
QY	856	TTAAACGTTATCTTTGAGTTTCGTCGTTTTAGGTGTTTCATCATGAGTAACTACAGTGA 915
Db	112618	TTAAACGTTATCTTTGAGTTTCATTTGTTTCAGGTTTCATTTAGTGAACCTACATGA 112677
QY	916	AAATCCTCTATGGACCCCTGATTTAATGTGGATGAAGCTAAGTCTCTGTCCTACTAGACCG 975
Db	112678	GAATCCTCAATGGATCCCTGATTTAATGATTAACGAAAGCTGAGTCTTGCT -CACTAGACCG 112736
QY	976	GAGTGAGAGCAACATGTTTACGAGAGCTATAGGATGAATTTGACGCTCTGACACTCGA 1035
Db	112737	GAGAGAGCAACATGTTTACGAGAGCTTCAAGGATGAATTTGAACGCTCCGACACTCGA 112796
QY	1036	CGTAAATCAAGAAGAGCTGAATCCCTAGAGAAAGAGGGCGATCTCGAGTAGATATCAG 1095
Db	112797	CGTAAATCAGAGAAGAGCTGAAATCGCTATAGAAAGAGGTGATGTCGAGTAGATATCAG 112856
QY	1096	CTGATTGATGAGGATATCAAACTGAGTATGAGCCAGAGTCATGGCCGAAGAGACGAAG 1155
Db	112857	CTGATTGATGAGGATATCGAACTGAGTATGAGCCAGAGTCATGGCAAGAGACGAAG 112916
QY	1156	CTACTGAAACAAATCCGACGAGGTTACGTTGAGAGAGTATATCAGATCTCTTTGAGTGAAT 1215
Db	112917	CTGCTGACAAACCCGACAGGTTACGTTGAGAGAGTATGTCAGATCTCTTTGAGATGAAC 112976
QY	1216	GACTTCTGGGAAACGAGGTATCCCTGATATGAGACTTTAGGCCAGTTGGGGTTACTGAG 1275
Db	112977	GACTTCTGGGAAACGAGGTATCCCTACTATGAGACTTTAGTCGAGCTGGGGTTACTGAG 113036
QY	1276	GACGTGACGATCTGTTTCGAGAAGTGTCATCTG ----- 1308
Db	113037	GACGTGACGATCAATTTCGAAGAGTGTCATCTGAGACACATCATGTCTTTACCCCTATGCT 113096
QY	1309	-----ATAAGGAGGAGACAATCGAGTTCTTTCCACCTGACCTGCAAGTGGAAATGATGAGGA 1364
Db	113097	GCCTATAAGGAGGAGACAATCGAGTTCTCTT-----ACTGCAAGTGGAGATGATGAGGGT 113152
QY	1365	CTCAGACTTTGAGCTGGATACCATGGGTTAGGGTTCTTGAAGTCTTTAGTGGATGAA 1424
Db	113153	CTCAGGACTTTGAGCTGGATACCATGGGTTAGGGTTCTTGAAGTCTCAGTGGATGAA 113212
QY	1425	CAGGGTACAGATTTAGATCAAGAAATGGAACAATGTTTGGTTTCCCTAGTGGAAAG 1484
Db	113213	CAGGGTATCAGCTTTTCATCAAGTAGTGGGAACAATGTTTGGTTTCACTAGTGGAAAG 113272
QY	1485	GGAAACCAACCCAGGTTTGACAGGAAGAGCTTTAAGGATTTGTGGCTACTATTGGGAAC 1544
Db	113273	GGAAACCAACCCAGGTTTGAACAGGAAGAGCTTTAAGGATTTGTGGGCTACTATTGGGAAC 113332
QY	1545	AATCTACCGCTAAACTCGACGCGTCCCAAGAGCAACCAATCCGGAGTCTCTGTGATTCGC 1604
Db	113333	AAT-TATCGCTAAACTCGGACGCTCCAAAGCAACCAATCCGGAGTCTCTGTGATTCGCG 113391
QY	1605	TACTTTTCAGGCTCGGTTGGCAATGTTTTTACTCCAGGAGTCTACAGGACCGGTGCT 1664
Db	113392	TACTTTTCAGTGTCTCGGTTGGCAACGTTTTCTACTCGAGGAGTCTCCAGGACCGGTGCT 113451
QY	1665	AACACGACATGAAGATGATAGATTCAGCGCTTATAGGGATTCCTCGCCTTACAAGGA 1724
Db	113452	AACAGGACATGAAGATGATAGATTCAGCGCTTATAGGGATTCCTCGCCTTACAAGGA 113511

[illegible]

ORIGIN

QY	2085	CCTGGCTTGAAGATCTTTATTTTCGAGGCGAGTCGCGCAACTGAGGAGATTAGTTCACACC	2144
DB	12077		
QY	2145	GAAGGAGCTACAATAGAAGATGTTGATGAGACATATGATATAGATGAGCGGAGTTTGAC	2204
DB	12017		
QY	2205	ACGAGCATGTATCTTTCAGTGAGCATATACCTCAGCGAGGAAAAGCAAGGTTTGAGC	2264
DB	11957		
QY	2265	GAAGCTCAGAGAAACAGCAAGCTGCGAAGTGGTGCAGGAGGAAAGCAAGGTTTGAGC	2324
DB	11897		
QY	2325	GCTAAGTGCCTCAGGCTATCAAGTTTCTGAAGGACAAGATCAGTGTCTCTTCCACT	2384
DB	11837		
QY	2385	ACAACTATCCGCAATGACAGTCCCTCAGGACATGCGCTTCGAGGAGATATGACGGCGCC	2444
DB	11777		
QY	2445	-----GAGCCTAGAGAGCAGAGATTCTGCATGTCCCTGCGAGG	2483
DB	11717		
QY	2484	CATTTCATCTTCGAGCTCGTGAATCTAGGAAGAATAGGAGACGACCTCACTCGATCT	2543
DB	11657		
QY	2544	AGCAGAGAGACAGACACTTCTGACGTCTCGTAGTTTACGCGACCGCGTCTGCGCGC	2603
DB	11598		
QY	2604	AATAGAAGAGAGAGTCCAGTATCTCTCAGAGCGGTCTGCGCGCCACAGAGCTGATGAG	2663
DB	11538		
QY	2664	ATCGAGTACCCATGCTGAGCTGATACGGAACATGGCGGGTTCGTCTATGGCTGGGAG	2723
DB	11478		
QY	2724	CANTCAGGACGACCTTACTACCACTTCGTTCTATTATTCGACTGAGGTAAGCGGCTC	2783
DB	11418		
QY	2784	ACTTCACCATATATATATCATCTCTTGTGATTGTTCTTTATTTTGTTCAGTGATTG	2843
DB	11358		
QY	2844	GATTGTCTGAGTACTCTCTTCCAGTTTATTCACACAGTGAGCTGTGATTAAAGTT	2903
DB	11298		
QY	2904	TGGGGAGGCGCTCAGGAA--GTATGTTGCAATGATATATATTTTAAAGTCTGCAATCATCT	2961
DB	11238		
QY	2962	AAGGCATAGAAAAACCAAAAAAATTTAAAAATTTTCAGAAATGATTTCACAAAAAGA	3021
DB	11178		
QY	3022	GTGTTTCAATGATGTCATTACATTTAGGATCAAGTCTAGAGTGTTTCAATTAGGATGTT	3081
DB	11119		
QY	3082	GCATATGCATAGGGGATAATCATGAGATAGCGCTTGTGAAGCA	3122
DB	11059	GCATGTGCATAGGGGATAATCATGAGATAGCGCTTGTGAAGCA	11019

RESULT 14

ATCHRIV9/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 53060)

Spiegel, L.A., Huang, E.N., Nascimento, L.U., de la Bastide, M., Vil, D.M., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A., Rodriguez, M., Shekher, M., Schutz, K., See, L.H., Swaby, I., Habermann, K., Dedhia, N.N., Mewes, H.W., Lemcke, K. and Mayer, K.F.X. Unpublished

2 (bases 21279 to 137726)

Zhong, J., Ma, P., Parnell, L.D., Chen, C.N., Chen, E.Y., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.

3 (bases 137729 to 137730)

Lamar, B., Stoneking, T., Stumpf, J., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.

4 (bases 1 to 159629)

EU Arabidopsis sequencing, project.

Direct Submission

Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemckemips.biochem.mpg.de.mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/> this fragment has an overlap with ATCHRIV8 at the 5' end and an overlap with ATCHRIV10 at the 3' end.

FEATURES

source

1..159629

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

6741..9588

/gene="AT4G03480"

complement(join(6741..7064,7161..7708,7783..8101,8409..8866,9283..9343,9445..9588))

/gene="AT4G03480"

complement(join(6741..7064,7161..7708,7783..8101,8409..8866,9283..9343,9445..9588))

/note="contains similarity to mammalian ankyrins similarity to similar to A. thaliana hypothetical protein F21B7.8, GenBank accession number AC002560 Contains ATP/GTP-binding site motif A (P-loop) AA297-304 contains EST gb:H37081, T42092"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAB77833.1"

/db_xref="GI:7270671"

/db_xref="SPTREMBL:Q9ZT76"

/translation="MVRSHLLRATPEASPLKSPASSPNSCRSLSPVELRLYFTRDDDF KKVVDLPKTSYQNSLPCLPCCSEFTLSRFDLNLPGEYVPMPEIIFSAMAGNI ELLEKLXSYETPMACLDGSDSVLHAAAGHLELVKNIITECPCLLEPNSKYQIP LHVAAAGRAAGVAVLVSVPSPVEDRDLRLVYKIDIDGDTPLHAALKDLHE KAEERINKLSHLIMHRRSRCISFSDSTRQNETAACLVADQHASFLANKDGTSP LYLAVEAGNVSILVRLMNFKNKIQGTSTLASOLEGRKSLHAAKAKNTDVLNVL NDDPSLVNRDEDEGRTCLSGASGYYKIGIKLDRSTKSYVECDKGSFPIHMAVEK GHLKVVKEILKRCPSDELNVKQGNMLHIAAKSAKVGSLGIRRLDTEHNLIEEQ

QY	976	GAGTGAGGCAACATGTTTACGAGAGCTATAGGAGTGAATTTGAACGCTCTGCACTCGA	1035	Db	149866	ATTCTTCTCCCTCGCAITGACGCTACTAGGAATCTCGAGGGAAGGAACATTTGACTTCAAG	149807
Db	150942	--GAGAGAGCAACATGTTTACGAGAGCTTACGAGAGTGAATTTGAATGTTCTGCACTCGA	150885	QY	2085	CCTGCGCTTGAAGATCTTTATTTTCGAGGGCAGTCCGCCAACTGAGGAGATTTAGTACACACC	2144
QY	1036	CGTAATCAAGAAGAGCTGAATTCGCTAGAGAAAGAGGGCGAGTCTGAGTAGATATGAG	1095	Db	149806	CCTGCGCTTGAAGATCTTTATTTTCGAGGGCAGTCCGCCAACTGATGATTTAGTACACACC	149747
Db	150884	CGTAATCAGAGAGTCTGTAATCCGCCAGAGAAAGAGGGCGATCTGAGTAGATATGAG	150825	QY	2145	GAAGGAGCTACAATAGAAAGATGTTGATGAGACATATGATATAGATGAGGGGAGTTTGAC	2204
QY	1096	CTGATTTGATGAGGATATCAAACTGAGTAGAGCCAGAGTCTATGGCGCAAGGAGACGAAG	1155	Db	149746	GAAGGAGCTCAACACAGAAAGATGTTGATGAGACAGATATATAGATGAGGGGAGTTCAAC	149687
Db	150824	CTGATTTGATGAGGATATCGAACTGAGTAGAGCAAGAGTCTATGGCGTAAGGAGACAAAG	150765	QY	2205	ACGAGCATGTTATCATTTTCAGTGAAGATATACCTCCAGCGAGGAAAGCAAGAGTTTGAGC	2264
QY	1156	CTACTGAACAATCCGACGAGGTTACAGTGGAGGAGTATATCAGATTTCTTTGAGATGAAT	1215	Db	149686	ACAAGCATGTTATCATTTTCAGTGAAGATATACCTCCAGCGAGGAAAGCAAGAGTTTGAGC	149627
Db	150764	CTGCGGAACAACCTGACGAGGTTACAGTGGAGGAGTATATCAGATTTCTTTAAGATGAAT	150705	QY	2265	GAAGCTCACAGAAACAAACAGCAAGCTGCAGAAAGTGGTCAAGAAACAGGATAAGTTACTC	2324
QY	1216	GACTTCTGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCAGTGGGGTTACTGGAG	1275	Db	149626	GAAGCTCACAGAAACAAACAGCAAGCTGCAGAAAGTGGTGCNAGAAACAGGATAAGTACTC	149567
Db	150704	GACTTCTGGGAAACGAGGTATTCGTGCTATGAGACTTTAGCCAGCTGGGGTTACTAGAG	150645	QY	2325	GCCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAAGATCAGCTGCTCTCTTCCACT	2384
QY	1276	GACGTGAGACTCTGTTTGAGAAAGTGTCTATCTG-----	1308	Db	149566	GTCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAAGATCAGCTGCTCTCTTCCACT	149507
Db	150644	GACGTGAGACTCTGTTTGAGAAAGTGTCTATCTGAGACACTCATGTTTACCCCTACGCT	150585	QY	2385	ACAACTATTTCCCAATGACAGCTCCCTCAGGACATGCTTCGAGGAGATATCAGCGCGCC	2444
QY	1309	----ATAAGGAGGAGACAATCGAGTTTCTTTCCACACTCAAGTGGAAATGTATGAGGA	1364	Db	149506	ACAGCCATTTCCGAGGGCAACTTTCTCAGGACATGCTTTTCGAGGAGATATGACGGCGCC	149447
Db	150584	GCTTATTTAGGAGGAGACAATCGAGTTTCTTTCCACACTCAAGTGGAGATGTATGAGGA	150525	QY	2445	-----GAGCCTAGAGACAGAAAGATTTCTGCATGTCCTCTCGAGG	2483
QY	1365	CTCACAGACTTTGAGCTGATACCAATGGGGTTAGGCTTCTTGACGTTCTTTAGTGGATGA	1424	Db	149446	GAGCCAAAGTCGGCGCAGGCTCAGCTAGAGAGCAAGAGATTCGCGATGCTCGCGAGG	149387
Db	150524	CTCACAGACTTTGAGCTGATACCAATGGGGTTAGGCTTCTTGACGTTCTTTAGTGGATGA	150465	QY	2484	CATTTCATCATTCGAGCTCGTGAATCTAGGAAGAATAGAGAACGACACTCTCATCGATCT	2543
QY	1425	CAGCGGTACAGATTTAGATCAAGAAATGGHAGACTGTTGGTTTCCCTAGTGGAAAG	1484	Db	149386	CATTTCATCATTCGAGCTCGTGAATCT--GGGAGGAGAGAGGCCACACTCATCTCATCT	149328
Db	150464	CAGCGGTATCAGCTTCGATCAAGAAATGGHAGACTGTTGGTTTCCCTAGTGGAAAG	150405	QY	2544	ACGACGAGAGCAGAGCAGCTTCTGAGCTCTGTTAGTTTACGCGACCGCGTGTGCGCGC	2603
QY	1485	GGAAACAAACCCAGGTTTGACAGGGAAGAGCTTAAGGATTTGTGGGCTACTATTGGGAAC	1544	Db	149327	AGCAGAGAGAGCAGAGCAGCTTCTACAGTCTCTGATGTTTACGCGACCGCGTGTGACCGC	149268
Db	150404	GGAAACAAACCCAGGTTTGACAGAGGAGCTTAAGGATTTGTGGGCTACTATTGGGAAC	150345	QY	2604	AATAGAAAGAGAGAGTCCGAGTATCCTCAGAGCGGTCTGGCGGCCACAGAGCTGATGAG	2663
QY	1545	AATCTACCGCTAAACTCGACGCGGTCCAAGAGCAACAAATCCGGAGTCTCTGTGATCGC	1604	Db	149267	AGTAGAAGAGAGAGTCCAGTATCCTCAGAGCGGTGCTGGCTGCCACAGAGCTGATGAG	149208
Db	150344	AATCTACCGCTAAACTTGCGCGGTCTAAGAGCAACAAATCCAAAGTCTCTGTGATCTGC	150285	QY	2664	ATCGAGTACCCACATGCTGGAGCTGATACGGAACATGGCGGTTCGTCTATGGCTCGGAG	2723
QY	1605	TACTTTACGGCTCGGTTGCCATGTTTCTTACTCCAGGAGCTACAGSCACCGTGTCT	1664	Db	149207	GTGAGTACCCACCTGCTGAGCTGATACGGAACAAGGGCGGTTCGTCTATGGCTCGGAG	149148
Db	150284	TACTTTACGGCTCGGTTGCCACGCTTTTACTCCAGGAGTCTACAGCACCGTGTCT	150225	QY	2724	CAATCACGGCAGCCATTGACTACCAACTTCGTTTCAATTTTCGACTGAGGTAAGCGCCTC	2783
QY	1665	ACACAGACATGAAGATGATAGATTACGCGCTTATAGGATTTCTCGCGCTTACAAAAGGA	1724	Db	149147	CAATCACATGACGCCATTGACGCCAACTCCGTTCAITCTTCGACTAAGGTAAGCGCCTC	149088
Db	150224	ACACAGACATGAAGATGATAGATTACGCGCTTACAGGATTTCTCGCGCTACGAAAGGA	150165	QY	2784	ACTTCACCAATTATATATATCATCTCTTGTGATTTGTTCTTTATTTTGTTCGATGATG	2843
QY	1725	AAGAATGCTCTGAGAGAGATCTTAACGATCTACACCAAGTATGCTCTGTTGATCCAT	1784	Db	149087	ACTTCACCAATTGATATATATCATCTCTTGTGATTTGTTCTTTATTTTGTTCGATGATG	149028
Db	150164	AAGAATGCTCTGAGAGCGATCTCAACAGCGCACCAAGGAAATG--TCTGTGATCCAC	150107	QY	2844	GATTTGCTGAGTACTCTCTTCCAAAGTTTATTCACACAGTGGACTGTGTGATTTAAGTT	2903
QY	1785	CTGTGGGTACATGAAGTGGCGCTGCACAAACGCAAGAGAGGTAAGAGAGGACTA	1844	Db	149027	GAITTTGCTGTTGTTACTCTCTTCCAAAGTTTATTCACACAGTGGTCTGTGTGATTTAAGTT	148968
Db	150106	CTGTGGGTATAGGAAGTGGCGGTGATGAACGGAAGAGAGGTTGAGAGGAGACTA	150047	QY	2904	TGGGGGAGGCTCAGGAA--GTATGTTGCAATGTATATATTTTAAAGTCTGCAATCATCT	2961
QY	1845	TGCGTGGGTGGGTGTGACGCCAAATCTGAAGTTTTGTGGAGTTTCGCTCAAGGAAGTA	1904	Db	148967	TAGGGGAGGCTCAGGAAAGTGTGTTGCAATGTATATATTTTGTGATCTGCAATCATCT	148908
Db	150046	TGCGTGGGTGGGTGTGATGCCAAATCTGGAAGTGTGGAAGTACCCTCAAGGAACCA	149987	QY	2962	AAGGCATAGAAAAACCAAAAAAATTTAAAAATTTCAAAAAATGATTTTCAAAAAAAGA	3021
QY	1905	GGGTTAGCACCGGAATGATGGAATTGATCACTTGGCGCGATGTGAGTTCTCTGAGTTT	1964	Db	148907	AAGGCATAGAAAAACCA--AAAAATTTAAAAAATTTAAGAAAAATGATTTCAAAAAAAGA	148849
Db	149986	GGTTTAGCACTGAGGATGATGGAATTTGCGGCAATTTGCGCGCATGTGAGTTCTTGGAGTTT	149927	QY	3022	GTGTTTCAATGATGTTGCAATTTAGGATCAAGTCTAGAGTGTTCATTTAGGATTTGTT	3081
QY	1965	GACATGTTGCGGACTTTCCCGCTACAGGTTTCAGGATTCATCATTTAGAAATCGCCAAC	2024	Db	148848	GTGTTTCAATGATGTTGCAATTTAGGATTTAGGATTTAGGATTTGTTTCAATTTAGGATTTGTT	148789
Db	149926	GACATGTTGCGGACTTTCCCGCTACAGGTTTCAGGATTCATCATTTAGAAATCGCCAAC	149867	QY	3082	GCATATGCATAGGGGATAAATGATGAGATAGCCTTTGTAAGCA	3122
QY	2025	ATTCTTTTCCCTGCAATTTACGCTACTAGGATTTCTCGAGGGCAGGAAACATTTCAAG	2084				

Db 148788 GCATGTGCATAGGCGATGAATGATGATAGCCTTGTAAACCA 148748

RESULT 15
AC074111/c
LOCUS
DEFINITION Arabidopsis thaliana chromosome 1 BAC T18N24 genomic sequence, complete sequence.
AC074111
VERSION AC074111.2 GI:22267605
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 98536)
AUTHORS Town,C.D., Haas,B.J., Maiti,R., Hannick,L.I., Chan,A.P., Ronning,C.M., Smith Jr.,R.K., Arbogast,T., Tallon,L.J., Uterback,T.R., VanAken,S.E., Feldblyum,T.V., Yu,C., Wortman,J.R., White,O. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC T18N24 genomic sequence
Unpublished
2 (bases 1 to 98536)
AUTHORS Town,C.D. and Kaul,S.
Direct Submission
Submitted (14-JUL-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
3 (bases 1 to 98536)
AUTHORS Town,C.D. and White,O.
Direct Submission
Submitted (16-AUG-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Aug 16, 2002 this sequence version replaced gi:9186844.
Address all correspondence to:at@tigr.org
BAC clone T18N24 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC clone.
Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://ccrmark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, <http://www.tigr.org/softlab/glimmer.htm> and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
Location/Qualifiers
1..98536
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="NULL"
/clone="T18N24"
61..2215
/gene="T18N24.40"
/note="contains similarity to auxin independent growth promoter (axi 1) GI:559921 from (Nicotiana tabacum)"
Join(<61..115,191..297,385..473,553..750,836..1040,
1144..1311,1390..1557,1639..1829,1984..2215)
/gene="T18N24.40"
CDS
product="growth regulator-related protein"
protein_id="AA094943.1"
/db_xref="GI:22267609"
/translation="MPKPCNSTVSSHFLNRLNGINVDIEPSESALAPSSGYLIVEA NGNLQQRSAICNAVAVAGLLNAVIVPFEFFHAIWKDSSNFGDIYDEHPIISLEGY VKIVRDPEDIMTRFSYNSVSIPIRVQAWATVNYNGEVYVILKEGVITRIFANR LAMSVPIYIQLLRCIANYKALKFSPISTLAELKLVDRMVEKSSATGGKYVSHLRFEE DMVAFSCLYBGGRAEKSEMDVIRQKSKGKFKERDRVIRPDILNRVNGKCLPTPLEVG MMLRGDFDNTSIVLAGRIYOPEKHLAPLOEMFRLYTKESLATPEELAPFOQYSS RMAALDYLSLSEVFTTQGGNFFHLMGHRRLFGHAKTIVIPDKPKVILLIQDME NMLSEVFKMKMLIGSDRGVMVPRVRKINRKTSTIYTPLECECIHLSNFSNTG NRSVGLALHPSSNLISSRL"
complement(2876..2897)
/rpt_family="AT_rich"
complement(3003..3070)
/rpt_family="(CAAC)n"
complement(4079..4172)
/rpt_family="AT_rich"
10087..14010
/gene="T18N24.39"
/notes="pseudogene, hypothetical protein"
/pseudo
complement(14695..14750)
/rpt_family="AT_rich"
15610..18756
/gene="T18N24.38"
/notes="pseudogene, hypothetical protein"
/pseudo
complement(16549..16704)
/rpt_family="(CATAA)n"
19928..25390
/gene="T18N24.37"
/notes="pseudogene, putative Athila ORF 1"
/pseudo
complement(25510..25533)
/rpt_family="(CA)n"
complement(28535..29290)
/gene="T18N24.44"
/note="contains similarity to transposon like ORF GI:6069596 from (Brassica rapa)"
complement(<28535..>29290)
/gene="T18N24.44"
complement(28535..29290)
/gene="T18N24.44"
/codon_start=1
/product="transposon-related protein"
protein_id="AA094941.1"
/db_xref="GI:22267607"
/translation="MTFANSHPIVSSRGKPHCPSCVKCKEKHIIISGRVSHLFSQG RNFEDYVWYKHGEMMDIRTDTYFTSTNHEVDNVVDPYMDVNDAPFNVVG YDNPFDHDDSYCNVEPENTENKRVYDLLEGANNPLDYDCRGQSLSLASRLMHKA EYINSEKLVNSVCEMTFDLPNGQATTSHTYQTEKLMRLNLGLPYHTIDICQNNCMILFW KEDEXEDQRCFCGAQRWRKPKDDRRSQIIMWVGF"
31586..35050
/gene="T18N24.36"
/note="pseudogene, hypothetical protein"
/pseudo
complement(39561..39594)
/rpt_family="AT_rich"
complement(40609..43080)
/gene="T18N24.43"
/note="pseudogene, putative Athila ORF 1"
/pseudo
complement(47106..48956)
/gene="T18N24.42"
/note="pseudogene, hypothetical protein"
/pseudo
FEATURES
source

Db 38800 TGTGGCTACTATTGGGAACAATTTACCGCTAAACTCGCGCGGTCCCAAGACAAACAAA 38741
QY 1585 TCCGAGTCCCTGTGATTCGCTACTTCAGCGCTCGGTTGCCAATGTTTTTACTCCAGGG 1644
Db 38740 TCCGAGTCCCTGGGATCCGCTACTTTCCCGCTCGGTTGCCAAGTTTTTACTCCAGGG 38681
QY 1645 AGTCTACAGGCACCGTGTCTAACACAGACATGAAGATGATGATTCAGGCTTATAGGGA 1704
Db 38680 AGTCTACAGGCACCGTGTCTAACACAGACATGAAGATGATGATTCAGGCTTATAGGGA 38621
QY 1705 TTCTCCGCTCTACAAAAGGAAGAAATGCTCTGAGAGGAGATCTTAAACGACTCACACCAG 1764
Db 38620 TTCTCCG-CTTACAAAGGAAGAAATGCTCTGAGAGGCGATCAACACGACCCAG 38562
QY 1765 TAATGCTCTGTTGATCCATCTGTGTGGGTACATGAAGTGGCGCTGACAAACGCGAAGA 1824
Db 38561 TAATGCTCTGTTATCCACTGTGTGATACAGGAAGTGGCAATTGACAAACGCGAAGA 38502
QY 1825 AGAAGGTAAAGGAGGACCTATCGGTGGGTGGCTGTGACGCCAATTCGAAAGTTGTG 1884
Db 38501 AGAAGGTGAGAGGAGCCTATGCGGTGGGTGGCTGTGACGCCAATTCGAAAGTTGTG 38442
QY 1885 GAGTTCGCTCAAGGAAGTAGGGTTAGCACCGAGAAATGATGACTTGGATCACTTCGCGC 1944
Db 38441 GAGTACCGCTCAAGGAGCCAGGTTTAGCACCGAGATGATGACTTGGATCACTTCGCGC 38382
QY 1945 GATGTGATCTCTGAGTTGACATGTTGGCGACTTTACCGCTACAGGTTTCGAGCAT 2004
Db 38381 GATGTGATCTCTGAGTTGATGACATGTTGGCGACTTTCCAGCGCTACAGGTTTCGAGCAT 38322
QY 2005 CATCGATTAGAAATCGCCAAATCTTTTCCCTCCATTTAGCTTACGTTAGGATTCGAGG 2064
Db 38321 CATCGATTAGAAATCGCCAAATCTTTTCCCTCCATTTAGCTTACGTTAGGATTCGAGG 38262
QY 2065 GCGAGAACATGACTTCAAGCCTCGCTTGAAGATCTTTATTCGAGGCGAGTCCGCCAA 2124
Db 38261 GAAGAACATGACTTCAAGCCTCGCTTGAAGATCTTTATTCGAGGCGGTGTCACCAA 38202
QY 2125 CTGAGGAGATTAGTCAACCGAAGGAGCTCAATAGAGATGTTGATGAGACATATGATA 2184
Db 38201 CTGAGGAGATTAGTCAACCGAAGGAGTTACACAGAGATGTTGATGAGACATATGATA 38142
QY 2185 TAGATGGCGGAGTTTGACAGAGCATGTATCATTTTCAGTGAGCATATACCTCCAGCGA 2244
Db 38141 TTGATGAGGAGAGATTTGACAGAGCATGTATCATTTTCAGTGAGATATACCTCCAGCGA 38082
QY 2245 GGAAAGCAAGAGTTTGAGCGAAGCTCACAGGACACACAGAGCTGCGAGAGTGGTGA 2304
Db 38081 GGAAAGCAAGAGTTTGAGCGAAGCTCACAGGACACACAAAGTTCAGAGTGGTGA 38022
QY 2305 AGAAACAGGATAGTTTACTCGCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAGA 2364
Db 38021 AGAAACAGGATAGTTTACTCGCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACTAGA 37962
QY 2365 TCAGCTGCTCCTCTCCACTACACTATTTCCGCAATGACAGCTCCCTCAGGACATGCCCT 2424
Db 37961 TCAGCTGCTCCTCTCCACTACACTATTTCCGCAAGTGCCTCAGGACATGCCCT 37902
QY 2425 CGAGGAGATATGACG-----GCCGAGCCTAGAGGACAGAGA 2463
Db 37901 CGAGGAGATATGACGCGCTGAGCCAGTCCGCGCGGCTGAGCCTAGAGGAGGAGA 37842
QY 2464 TTCTGATGTCCTCGGAGGCAATTCATTCAGGCTCGTGAATCTAGGAAGATAGGA 2523
Db 37841 TTCCGATGTCCTCGGAGGCAATTCATTCAGGCTCGTGAATCTAGGAAGATAGGA 37782
QY 2524 GAACGACACTCACTCGATCTAGCAGGAGGAGAGAGCTTCTGCGAGTCTCGTAGTTTAC 2583
Db 37781 GGACCGAGCTCACTCGATCTAGCAGGAGGAGAGAGCTTCTGCGAGTCTCGTAGTTTAC 37722
QY 2584 GCGACCGCGTGTGCGCGCAATAGAAAGAGAGGTCGAGTATCCTCAGAGCGGTGCTG 2643
Db 37721 GCGACCGCGTGTGCGCGCAATAGAAAGAGAGGTCGAGTATCCTCAGAGCGGTACTG 37662

QY 2644 GCCGCCACAGAGCTGATGAGATCGAGTACCCACATGCTGGAGCTGATACGGAACATGGCG 2703
Db 37661 GCTGCCACAGAGCTGATGAGTCAAGTACCACTTGCCTGGAGATGATACAGAAACAGGGCG 37602
QY 2704 GTTCGTCTATGGCTTGGGAGCAATCACAGGAGCCATTGACTACCAACTTCGTTCAATTAT 2763
Db 37601 GCTCGTCTATGGCTGAGAGCAATCACAGGAGCCATTGAGCAACACTCCGTTAATTCT 37542
QY 2764 TCGACTGAGGTAAAGGCTCACTTCAACATTATATATATCATCTCTTGTGATTTGTTCT 2823
Db 37541 TCGACTAAGSTAAGCGGCTCACTTCAACCTGTATATATCATCTCTTGTGATTTGTTCT 37482
QY 2824 TTATTTTGTCTTCACTGATTTGATTTGCTCAGTACTCTCTTCCAAAGTTTATTCACACAG 2883
Db 37481 TTATTTTGTCTTCACTGATTTGATTTGCTCAGTACTCTCTTCCAAAGTTTATTCACACAA 37422
QY 2884 TGGACTGTGTGATTTAAAGTTTGGGGAGGGCTCAGGAA--GTATGTTGCACTTGTATATAT 2941
Db 37421 TGGACTGTGTGATTTAAAGTTTGGGGAGGGCTCATGAAGTGTGTTGTCATTGTATATAT 37362
QY 2942 TTTTAACTCTGCATTCATCTAAAGCATAGAAAACCAAAAATAAATAAATTTCAAA 3001
Db 37361 TCTTGAAGTCTACATTCATCTAAAGCATAGTAAAAACAAAAAATTTCAAAAATTTCAAAA 37302
QY 3002 AATGATTTCAAAAAAAGAGTGTTCATGTAGTTGCAATTACATTTAGGATCAAGTCTAGA 3061
Db 37301 AATGATTTCAAAAAACAGAGTGTTCATGTAGTT--CATTTACATTTAGGATCGAGTCTAGA 37243
QY 3062 GTGTTTCATTTAGGATTTGTCATATGCATAGGGGATAATGATGAGATAGCCTTTGTAAGC 3121
Db 37242 GTGTTTCATTTAGGATTTGTCATATGCATAGGGGATAATGAGATAGCCTTTGTAAGC 37183
QY 3122 A 3122
Db 37182 A 37182

Search completed: September 9, 2004, 07:15:41
Job time : 11679 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2004, 02:31:27 ; Search time 1106 Seconds
(without alignments)
11991.749 Million cell updates/sec

Title: US-10-088-384A-27
Perfect score: 3122
Sequence: 1 ttctatatattgcactcttc.....atgagatagccttgtaagca 3122

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002s:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3122	100.0	3122	4	Aaf30459 Transcrip
2	1638.4	52.5	96988	3	Aaf22290 BAC conta
3	1594.6	51.1	110000	3	Continuation (10 o
4	1588.2	50.9	83390	3	Aaf22283 BAC conta
5	1588.2	50.9	90336	3	Aaf22289 BAC conta
6	1415.6	45.3	129021	3	Aaf22296 BAC conta
7	1398.6	44.8	4860	4	Aaf30458 Transcrip
8	1366.8	43.8	67087	3	Aaf22280 BAC conta
9	1339	42.9	64415	3	Aaf22279 BAC conta
10	1311	42.0	83390	3	Aaf22283 BAC conta
11	1311	42.0	90336	3	Aaf22289 BAC conta
12	1230.8	39.4	2512	4	Aaf30450 Transcrip
13	1215.4	38.9	92584	3	Aaf22288 BAC conta
14	1174.8	37.6	96988	3	Aaf22290 BAC conta
15	1127.2	36.1	1997	4	Aaf30451 Transcrip
16	1108.6	35.5	134499	3	Aaf22286 BAC conta
17	1066.2	34.2	163319	3	Aaf22306 Arabidopsi
18	1044.6	33.5	129021	3	Aaf22296 BAC conta
19	1018.8	32.6	79122	3	Aaf22294 BAC conta
20	1002.6	32.1	59590	3	Aaf22281 BAC conta
21	988.2	31.7	86584	3	Aaf22292 BAC conta
22	981.8	31.4	82588	3	Aaf22301 BAC conta
23	955.6	30.6	110000	3	Continuation (9 of

24	751.6	24.1	1857	3	AAZ35276 Arabidops
25	513.2	16.4	163319	3	Aaf22306 Arabidops
26	469	15.0	95223	3	Aaf22282 BAC conta
27	467.8	15.0	109973	3	Aaf22298 BAC conta
28	449	14.4	40349	3	Aaf22278
29	413.6	13.2	110000	3	Aaf22303 Arabidops
30	409.4	13.1	82138	3	Continuation (11 o
31	409.4	13.1	94895	3	Aaf22302 BAC conta
32	408.6	13.1	103929	3	Aaf22287
33	408.6	13.1	134499	3	Aaf22286
34	358.6	11.5	95223	3	Aaf22282
35	357.6	11.5	86584	3	Aaf22292
36	343.6	11.0	64415	3	Aaf22279
37	217.4	7.0	80450	3	Aaf22295
38	214	6.9	903	4	Aaf30454 Transcrip
39	180.2	5.8	40349	3	Aaf22278
40	160.4	5.1	96583	3	Aaf22297
41	137	4.4	82138	3	Continuation (11 o
42	137	4.4	82588	3	Aaf22301
43	137	4.4	94895	3	Aaf22302
44	137	4.4	110000	3	Continuation (10 o
45	122.8	3.9	110000	3	Aaf22303 Arabidops

ALIGNMENTS

RESULT 1
AAF30459
ID AAF30459 standard; DNA; 3122 BP.
XX
AC AAF30459;
XX
DT 29-MAY-2001 (first entry)
XX
DE Transcriptionally silent information clone TSI-A genomic DNA.
XX
KW TSI-A; transcriptionally silent information; gene silencing; ds.
XX Arabidopsis thaliana.
XX
PN WO200120010-A2.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-EF008994.
XX
PR 16-SEP-1999; 99GB-00021964.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PA Steiner A, Mittelsten Scheid O, Paszkowski J;
XX
PI WPI; 2001-244815/25.
XX
PT Novel transcriptionally silenced genes obtained from Arabidopsis useful
for selecting plants which compared to wild type plant are impaired in
transcriptional gene silencing.
XX
PS Claim 1; Page 34-35; 36pp; English.
XX
CC This present sequence is that of a transcribed region of Arabidopsis
thaliana genomic DNA identical to nucleotides 65081-68202 of BAC F7N22.
CC The sequence was isolated from an Arabidopsis genomic DNA database on the
basis of homology to Arabidopsis transcriptionally silent information
(TSI) sequences. It is located 5' to the cDNA region given in AAF30456.
CC Comparison of transcriptional gene expression between an Arabidopsis line
carrying a silent transgene present in multiple copies and its mutant
derivative moml impaired in silencing of the transgene revealed 2 cDNA
clones which were expressed in the mutant plants, but not in the parental
or wild-type plants. Both clones were derived from the same family of
transcripts referred to as TSI. Genomic templates encoding TSI are
repetitive elements with mainly pericentromeric location and conserved

CC organisation among various ecotypes. Transcriptional silencing of genomic
 CC TSI templates is specifically released in the mutant. Transcription of
 CC a plant. The invention provides a nucleic acid encoding silencing pathway in
 CC information which is silencing in wild-type plants but whose expression is
 CC turned on in plants defective in transcriptional gene silencing, and has
 CC the formula: RA-RB-RC, where RA, RB, RC constitute component nucleotide
 CC sequences selected from the present sequence and those given in AAF30450-
 CC 58. Also provided are methods of selecting a plant that is impaired in
 CC transcriptional gene silencing, and of producing DNA representing at
 CC least part of a gene necessary to maintain silencing of another gene in a
 CC cell or plant

XX Sequence 3122 BP; 855 A; 767 C; 691 G; 809 T; 0 U; 0 Other;

Query Match	100.0%;	Score 3122;	DB 4;	Length 3122;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 3122;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	TTTCATATATTCGACCTCTTCTCTCATCTTCTGATCCAAAGACACAAACGCGCCATC	60		
DB	1	TTTCATATATTCGACCTCTTCTCTCATCTTCTGATCCAAAGACACAAACGCGCCATC	60		
QY	61	GCTTTCCCTCAACAACCTCTCACTCGACCAACGCGCGCTCTCTCACTTACTCGGCTTCATCG	120		
DB	61	GCTTTCCCTCAACAACCTCTCACTCGACCAACGCGCGCTCTCTCACTTACTCGGCTTCATCG	120		
QY	121	CTCTCATCGCATCTCTCAACATCTGACATCTGACCTCGGATATCACTCGAGCTCCGCTCTCT	180		
DB	121	CTCTCATCGCATCTCTCAACATCTGACATCTGACCTCGGATATCACTCGAGCTCCGCTCTCT	180		
QY	181	CACCGCTCTCATCTGCTCAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	240		
DB	181	CACCGCTCTCTCATCTGCTCAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	240		
QY	241	ATTTCATCTCACTCGACCTCTTACCAACGCGCTCTCTCACTCTCTCTCTCTCTCTCTCTCT	300		
DB	241	ATTTCATCTCACTCGACCTCTTACCAACGCGCTCTCTCACTCTCTCTCTCTCTCTCTCTCT	300		
QY	301	CGACCTCTCACTCAACGCT	360		
DB	301	CGACCTCTCACTCAACGCT	360		
QY	361	CGACCTCTCACTCAACGCT	420		
DB	361	CGACCTCTCACTCAACGCT	420		
QY	421	GCTCTCATCT	480		
DB	421	GCTCTCATCT	480		
QY	481	AGCAACCAACTCGACCT	540		
DB	481	AGCAACCAACTCGACCT	540		
QY	541	TCCTTCATCT	600		
DB	541	TCCTTCATCT	600		
QY	601	CTCACTCGATTTGTCAG	660		
DB	601	CTCACTCGATTTGTCAG	660		
QY	661	CCCTCTACATCTCAACCTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	720		
DB	661	CCCTCTACATCTCAACCTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	720		
QY	721	CATACCTCGACCACT	780		
DB	721	CATACCTCGACCACT	780		
QY	781	CAGTTTCACTCAACCGGTTTACTCGACCGGTACGCTGGTTAGATTGTTTGGTTT	840		

DB	781	CAGTTTCACTCAACCGGTTTACTCGACCGGTACGCTGGTTAGATTGTTTGGTTT	840		
QY	841	GCTATTACTAATATTAACGTTTATCTTTGAGTTTCTGTTGTTTGGTTTGGTTTGGTTT	900		
DB	841	GCTATTACTAATATTAACGTTTATCTTTGAGTTTCTGTTGTTTGGTTTGGTTTGGTTT	900		
QY	901	AGTAACCTACAGTGGAAATCTCTATGACCCCTCATTAATATGATGGAAGAACTAAGTCC	960		
DB	901	AGTAACCTACAGTGGAAATCTCTATGACCCCTCATTAATATGATGGAAGAACTAAGTCC	960		
QY	961	TGTTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATPAGGATGAATTTGAA	1020		
DB	961	TGTTCCACTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATPAGGATGAATTTGAA	1020		
QY	1021	CGCTCTGAGCTCGACGTAATCAAAAGAGCTGAAATCGCTAGAGGAAGAGGCGGATG	1080		
DB	1021	CGCTCTGAGCTCGACGTAATCAAAAGAGCTGAAATCGCTAGAGGAAGAGGCGGATG	1080		
QY	1081	TCGAGTAGATATGAGCTGATGATGAGGATATCAAACTGAGTATGAGCCAGGCTATGG	1140		
DB	1081	TCGAGTAGATATGAGCTGATGATGAGGATATCAAACTGAGTATGAGCCAGGCTATGG	1140		
QY	1141	CGCAAGGAGAGCAAGCTCTCTGAAACAAATCCGACGAGGTTACAGTGGAGGATATATCAG	1200		
DB	1141	CGCAAGGAGAGCAAGCTCTCTGAAACAAATCCGACGAGGTTACAGTGGAGGATATATCAG	1200		
QY	1201	TTCTTTGAGATGAATGACTTCTGGGGAACGAGGATCTCTGATATGAGACTTTAGCCAG	1260		
DB	1201	TTCTTTGAGATGAATGACTTCTGGGGAACGAGGATCTCTGATATGAGACTTTAGCCAG	1260		
QY	1261	TTGGGGTTACTCGAGGAGCTGAGCATCTGTTTCGAGAACTGTCTATCTGTAAGGAGGAG	1320		
DB	1261	TTGGGGTTACTCGAGGAGCTGAGCATCTGTTTCGAGAACTGTCTATCTGTAAGGAGGAG	1320		
QY	1321	CAATCGAGTTTCTTTTCCACACTGCAAGTGGAAATGTATGAGGACTCTACAGACTTTGAG	1380		
DB	1321	CAATCGAGTTTCTTTTCCACACTGCAAGTGGAAATGTATGAGGACTCTACAGACTTTGAG	1380		
QY	1381	TGATATACCAATGGGTTAGGCTTCTGACGTTCTTAGTGGATGAGGAGGCTACAGATTT	1440		
DB	1381	TGATATACCAATGGGTTAGGCTTCTGACGTTCTTAGTGGATGAGGAGGCTACAGATTT	1440		
QY	1441	AGATCAAGAAATTTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGGT	1500		
DB	1441	AGATCAAGAAATTTGGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAAACCAACCCAGGT	1500		
QY	1501	TTGACAGGAGAGGACTTTAAGGATTTGTGGCTACTATTTGGGAAACAATCTACCGCTAACT	1560		
DB	1501	TTGACAGGAGAGGACTTTAAGGATTTGTGGCTACTATTTGGGAAACAATCTACCGCTAACT	1560		
QY	1561	CGACCGGCTCAAGAGAGCAACCAATCGGAGTCTGATTTGCTGATTTCCGCTCTTTAGCGCTCGG	1620		
DB	1561	CGACCGGCTCAAGAGAGCAACCAATCGGAGTCTGATTTGCTGATTTCCGCTCTTTAGCGCTCGG	1620		
QY	1621	TTGCCAATGTTTTTACTCCAGGAGTCTACAGGACCGGTCTGATTTCCGCTCTTTAGCGCTCGG	1680		
DB	1621	TTGCCAATGTTTTTACTCCAGGAGTCTACAGGACCGGTCTGATTTCCGCTCTTTAGCGCTCGG	1680		
QY	1681	TGATAGATTTCAGCGCTTTATAGGATTTCTCCGCTTTACAAAGGAAAGAAATGCTCTGAGAG	1740		
DB	1681	TGATAGATTTCAGCGCTTTATAGGATTTCTCCGCTTTACAAAGGAAAGAAATGCTCTGAGAG	1740		
QY	1741	GAGATCTTAACGACTCACACCGAGTATGCTGTTGATCTGATCTGTTGGGTACATGA	1800		
DB	1741	GAGATCTTAACGACTCACACCGAGTATGCTGTTGATCTGATCTGTTGGGTACATGA	1800		
QY	1801	AGTGGCGCTGACAAACGCGCAAGAGAGGTTAAGGAGGACTATGCTGGGTGGCGTTG	1860		
DB	1801	AGTGGCGCTGACAAACGCGCAAGAGAGGTTAAGGAGGACTATGCTGGGTGGCGTTG	1860		
QY	1861	TGACGCCAAATCTGAAAGTTTGTGGAGTTCCGCTCAAGGAGAGTGGGTAGCACCGAGAA	1920		
DB	1861	TGACGCCAAATCTGAAAGTTTGTGGAGTTCCGCTCAAGGAGAGTGGGTAGCACCGAGAA	1920		

```
QY 1921 TGATGGACTTGGATCACTTGGCCGATGTGAGTTCTCTGAGTTTGACATGTTGGCGACT 1980
Db 1921 TGATGGACTTGGATCACTTGGCCGATGTGAGTTCTCTGAGTTTGACATGTTGGCGACT 1980
QY 1981 TTCAACCGCTACAGGTTTGAGGATTCATCGAATTAGAAATCGCCAACTTTTCCCTCGCA 2040
Db 1981 TTCAACCGCTACAGGTTTGAGGATTCATCGAATTAGAAATCGCCAACTTTTCCCTCGCA 2040
QY 2041 TTTACGCTACTAGGATTTCTGAGGCGAGGAACATTTGACTTCAAGCCTGCGCTCAAGATC 2100
Db 2041 TTTACGCTACTAGGATTTCTGAGGCGAGGAACATTTGACTTCAAGCCTGCGCTCAAGATC 2100
QY 2101 TTTATTTGAGGGCAGTCCGCCAACTCGAGGATTTAGTTCACACCGAAGAGCTACAATAG 2160
Db 2101 TTTATTTGAGGGCAGTCCGCCAACTCGAGGATTTAGTTCACACCGAAGAGCTACAATAG 2160
QY 2161 AAGATGTTGATGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCAAT 2220
Db 2161 AAGATGTTGATGACATATGATATAGATGAGGCGGAGTTTGACACGAGCATGTATCAAT 2220
QY 2221 TCAGTGAAGCATATACCTCCAGCGAGGAAGCAAGAGTTTGGAGGAAGCTCACAGGAACA 2280
Db 2221 TCAGTGAAGCATATACCTCCAGCGAGGAAGCAAGAGTTTGGAGGAAGCTCACAGGAACA 2280
QY 2281 ACAGCAAGCTGCAGAACTGTTGCAAGAAACAGGATAAGTTTACTCGCCCAAGTGCCTCAGGG 2340
Db 2281 ACAGCAAGCTGCAGAACTGTTGCAAGAAACAGGATAAGTTTACTCGCCCAAGTGCCTCAGGG 2340
QY 2341 CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCTCTCTTCCACTACAACTATTCCGCAAT 2400
Db 2341 CTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCTCTCTTCCACTACAACTATTCCGCAAT 2400
QY 2401 GACAGCTCCCTCAGGACATGCTTCGAGGAGATGATGACGCGCCGAGCCTAGAGAGAGA 2460
Db 2401 GACAGCTCCCTCAGGACATGCTTCGAGGAGATGATGACGCGCCGAGCCTAGAGAGAGA 2460
QY 2461 AGATTTCTGATGTCCTCGAGGCAATTCATTCGAGGCTCGTGAATCTAGGAAGATA 2520
Db 2461 AGATTTCTGATGTCCTCGAGGCAATTCATTCGAGGCTCGTGAATCTAGGAAGATA 2520
QY 2521 GGAGAGGACACTCACTCGATCTAGCAGGAGGAGCAGCACTTCTGAGTCTCGTAGTT 2580
Db 2521 GGAGAGGACACTCACTCGATCTAGCAGGAGGAGCAGCACTTCTGAGTCTCGTAGTT 2580
QY 2581 TAGCGACCGCGTGTGCGGCAATAGAAAGAGAGGTGCGAGTATCCTCAGAGCGGTG 2640
Db 2581 TAGCGACCGCGTGTGCGGCAATAGAAAGAGAGGTGCGAGTATCCTCAGAGCGGTG 2640
QY 2641 CTGGCGCCACAGAGCTGATGAGATCGAGTACCACATGCTGGAGCTGATACGGAACATG 2700
Db 2641 CTGGCGCCACAGAGCTGATGAGATCGAGTACCACATGCTGGAGCTGATACGGAACATG 2700
QY 2701 GCGGTTCTGATGCTTGGGCAATCAGCAGGCAATGACTACCACTTCGTTTCAAT 2760
Db 2701 GCGGTTCTGATGCTTGGGCAATCAGCAGGCAATGACTACCACTTCGTTTCAAT 2760
QY 2761 TATTCGACTGAGGTAAAGCGCTCACTTCACTTATATATATATATATATATATATATATAT 2820
Db 2761 TATTCGACTGAGGTAAAGCGCTCACTTCACTTATATATATATATATATATATATATATAT 2820
QY 2821 TCTTTATTTGTTTTCAGTATGATTTGTCCTGAGTACTCTCTTCCAAAGTTTATTCACA 2880
Db 2821 TCTTTATTTGTTTTCAGTATGATTTGTCCTGAGTACTCTCTTCCAAAGTTTATTCACA 2880
QY 2881 CAGTGACTGTGTGATTTAAGTTTGGGGGAGGCTCAGGAAGTATGTTGCAATGTATATA 2940
Db 2881 CAGTGACTGTGTGATTTAAGTTTGGGGGAGGCTCAGGAAGTATGTTGCAATGTATATA 2940
QY 2941 TTTTAAAGTCTGATTCATCTAAGGATAGAAAACCAAAAAAATTTAAAAATTTTTCAGA 3000
Db 2941 TTTTAAAGTCTGATTCATCTAAGGATAGAAAACCAAAAAAATTTAAAAATTTTTCAGA 3000
```

```
QY 3001 AAATGATTTCAAAAAAAGAGTGTTCATGTAGTTGCAATTACATTTAGGATCAAGTCTAG 3060
Db 3001 AAATGATTTCAAAAAAAGAGTGTTCATGTAGTTGCAATTACATTTAGGATCAAGTCTAG 3060
QY 3061 AGTGTTCATTTAGGATTTGCAATGATGATGAGGATAATGATGAGATAGCCTTTGTAAG 3120
Db 3061 AGTGTTCATTTAGGATTTGCAATGATGATGAGGATAATGATGAGATAGCCTTTGTAAG 3120
QY 3121 CA 3122
Db 3121 CA 3122
```

RESULT 2

AAF22290 ID AAF22290 standard; DNA; 96988 BP.

XX AAF22290;

XX AC 20-MAR-2001 (first entry)

XX DE BAC containing repeats from centromeres 1-4 #13.

XX KW Centromere; michromosome; vector; ds.

XX OS Arabidopsis thaliana.

XX PN W020005325-A2.

XX XX 21-SEP-2000.

XX PF 17-MAR-2000; 2000WO-US007392.

XX PR 18-MAR-1999; 99US-0125219P.

XX PR 01-APR-1999; 99US-0127409P.

XX PR 18-MAY-1999; 99US-0134770P.

XX PR 13-SEP-1999; 99US-0153584P.

XX PR 17-SEP-1999; 99US-0154603P.

XX PR 16-DEC-1999; 99US-0172493P.

XX XX (UYCH-) UNIV CHICAGO.

XX PI Preuss D, Copenhaver G, Keith K;

XX DR WPI; 2000-587529/55.

XX XX Recombinant DNA construct comprising a plant centromere, useful for

XX PT producing stably inherited michrosomes which can serve as vectors for the

XX PT construction of transgenic plant and animal cells.

XX PS Claim 102; Page 550-572; 1449pp; English.

XX XX The present invention relates to a recombinant DNA construct of a plant

XX CC (Arabidopsis thaliana) centromere. The constructs are useful for

XX CC producing stably inherited michrosomes which can serve as vectors for the

XX CC construction of transgenic plant and animal cells expressing selected

XX CC proteins such as hormones, enzymes, interleukins, clotting factors,

XX CC cytokines, antibodies, and growth factors

XX SQ Sequence 96988 BP; 29654 A; 18289 C; 18523 G; 30522 T; 0 U; 0 Other;

Query Match 52.5%; Score 1638.4; DB 3; Length 96988;

Best Local Similarity 89.6%; Pred. No. 0;

Matches 1874; Conservative 0; Mismatches 156; Indels 62; Gaps 8;

QY 1090 TATGAGCTGATTCAGGATATCAAACTGAGTATGAGCAGAGTCATGGCGCAGGAG 1149

Db 54120 TAAACCTTATTATCAAGGACATAGAACTGAGTATGAGCAGAGTCATGGCTTAAGGAG 54179

QY 1150 AGAAGCTTACTGAAACAAATCCGACGAGGTTACAGTGGAGGATATATCAGATTCCTTTGAG 1209

Db 54180 ACAAAGCTGTGAAACAAACCCGACGAGGTTACAGTGGAGGATATATCAGATTCCTTTGAG 54239

QY 1210 ATGAATGACTTCTGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCCACTTGGGGTTA 1269
Db |||||
QY 54240 ATGAATGACTTCTGGGAAACGAGGTATCCCTGATGAGAAATTTAGCCCACTTGGGGTTA 54299
Db |||||
QY 1270 CTGAGGAGCGTGCAGCTCTCTTTCGAGAAAGTGTCACTG----- 1308
Db |||||
QY 54300 CTGAGGAGCGTGCAGCTCTCTTTCGAGAAAGTGTCACTGAGACACTCATGCTCTTACCCCT 54359
QY 1309 -----ATPAGGAGGAGCAATCGAGTTCTTTCCACACTGCAAGTGAAGATGAT 1358
Db |||||
QY 54360 TACCTTGCCTATAGGAAAGACAAATCGAGTTCTTTTTCACGCTGCAAGTGAAGATGAT 54419
QY 1359 GAGGAGCTCACAGACTTTGAGCTGGATACCACTGGGGTTAGGCTTCTTGACGCTTCTTAGTG 1418
Db |||||
QY 54420 AAGGCACTCAAAACTTTAGCTGATACCACTGGGGTTAGGCTTCTTGACGCTTCTCACTG 54479
Db |||||
QY 1419 GATGAACAGCGGTACCAATTTAGATTAAGAAATTTGGAAGAACTCTTTGGTTTCCCTAGT 1478
Db |||||
QY 54480 AATGAACAGCGGTATCAGCTTTCGATCAACAAAGTTGGAAGAACTCTTTGGTTTCCCTAGT 54539
QY 1479 GGAAGGGAACCAACCCAGGTTTCACAGGGAAGAGCTTAAGGATTTGTGGGCTACTATT 1538
Db |||||
QY 54540 GGAAGGGAACCAACCCAGGTTTCACAGAGAAAAGCTTAAGGATTTGTGGGCCATTATA 54599
QY 1539 GGAACCAATCTACCGCTAAACTCGACGGGTTCCAAAGAGCAACCAATCCGGAGTCTCTGTG 1598
Db |||||
QY 54600 GGAAC-ATCTACCGCTAAACTCGGTTGGTCCAGAGCAACCAATCCGGAGTCTCTGTG 54658
QY 1599 ATTCGCTACTTTACGCGTTCGTTGCCAATGTTTTTACTCCAGGAGTCTACAGGCACC 1658
Db |||||
QY 54659 ATCTGCTACTTTACAGAGATGTTGCCAAGCTCTTTTACTCCAGGAGTCTACAGGCACC 54718
QY 1659 GTGCTCAACAGACATGAAGATGATAGATTCAGCGCTTATAGGATTTCTCGCCTTACA 1718
Db |||||
QY 54719 GTGCTCAACAGGACATGAGATGATAGATTTAGCGCTTACAGGATTTCTCGCGGAGC 54778
QY 1719 AAAGGAAAGAAATGCTCTGAGAGGAGATCTTAACGATCTCAACCAAGTAATGCTCTGTG 1778
Db |||||
QY 54779 AAGGGAAGAAATGCTTGAAGCGGATCTCAACAGCGCAACCAAGTAATGCTCTGTG 54838
QY 1779 ATCCATCTGTGGGTACATGAAGTGGCGCTGCAACAGCGCAAGAGAGGTAGAGGA 1838
Db |||||
QY 54839 ATCCATCTGTGGGTACAGGAAGTGGCGGTTGACGAACGGAAGAGAGGTAGAGGA 54898
QY 1839 GCATATGCTGGGTGGGCTGTGAGCGCAATCTGAAAGTTTGTGGAGTTCGGCTCAAG 1898
Db |||||
QY 54899 GCACAAATGCTGGGTGGGCTGTGAGCGCAATCTGAAAGCTTGTGGAGTACCGCTCAAG 54958
QY 1899 GAAAGTGGGTTAGCACCGGAAATGATGGACTTGGATCACTTGGCGCGATGTGAGTTCTCT 1958
Db |||||
QY 54959 GAACCAAGGTTAGCACCGGAGGATGATGGACTTGGATCAATTTGGCGCAGATGTGAGTTCTTG 55018
QY 1959 GAGTTGACATGTTGGGCACTTTACCGCTACAGGTTTCAGGATTCATCGATTAAGAATC 2018
Db |||||
QY 55019 GAGTTTGACATGCTGGCAACTTTACCGCTACAGGTTTCAGGATTCATCGATTAAGAATC 55078
QY 2019 GCCAATCTTTTCCCTGCAATTTACGCTTACTAGGATTTCTGAGGCGCAGGAACATGAC 2078
Db |||||
QY 55079 GCCATCTTTCTCCCTGCAATTAAGCTACTAGGATTTCTGAGGCGCAGGAACATGAC 55138
QY 2079 TTCAAGCCTCGCTGAAGATCTTTATTTGAGGGGAGTCCGCCAATCTGAGGAGATGAT 2138
Db |||||
QY 55139 TTCAAGCCTCGCTTGAAGATCTTTATTTGAGGGTGTGTTCAACAACTGAGGAGATCAGT 55198
QY 2139 CACACCGAAGGAGCTCAATAGAGATGTTGATGACACATATCATATAGTATGAGGCGGAG 2198
Db |||||
QY 55199 CATACCGAAGGAGCTCAACAGAAAGATGTTGATAAGACGGATGATATAGATGAGGCGGAG 55258
QY 2199 TTTGACACGAGCTGATCATTTTCACTGAGCATATACCTCCAGCGAGGAAAAAGCAAGAT 2258
Db |||||
QY 55259 TTTGACACGAGCTGATCATTTTCACTGAGCATATACCTCTCGGAGGAAAAAGCAAGAT 55318
QY 2259 TTGAGCGAAGCTCACAGGAACAAACAGCAAGCTGCAAGAGTGGTGTCAAGAAACAGGATAAG 2318

Db 55319 TTGAGCGAAGCTCAAGAAC-ACAGCAAGCTGCAGAAAGTGGTGCAGAAACATGATTAAG 55377
QY 2319 TTACTCGCAGAGTCCCTCAGGGCTATCAAGTTTCTGAAGGACAAAGATCAGTGTCTCTCT 2378
Db 55378 TTACTCGCAGAGTCCCTCAGGGCTATCAAGTTTCTGAAGGACAAAGATCAGTGTCTCTCT 55437
QY 2379 TCCACTACAACTATTCCGCAATGACAGCTCCCTCAGGACATGCTTCCAGAGAGATATGAC 2438
Db 55438 TCCACTACAGCTATCCCGCAGGGACATCTTCTCAGGACATGCTTCCAGGAGATATGAT 55497
QY 2439 GCGCCC-----GAGCCTAGAGAGCAGAAAGATTCGATCTGCTCCCT 2477
Db 55498 GCGCCCGAGCCAAAGTCAAGCGCAGGCTGAGCCTAGAGAGCAGAGATTCGCAATGCTCCCT 55557
QY 2478 GCGAGGCAATTCATATTCGAGCCTCGTGAATCTTAGAAGAAATAGAGAAACGACACTACT 2537
Db 55558 GCGAGGCAATTCATATTCGAGCCTCGTGAATCTGAGGAAAGAGAGGAGCACAACACTACT 55617
QY 2538 CGATCTAGCAGAGAGCAGAGCTTCTGAGCTCTCGTAGTCTGAGCGACCGCGGTGCT 2597
Db 55618 CGATCTAGAGTAGGAGCAGAGCTTCTGAGTCTCGTAGTCTGAGCGACCGCGATGCT 55677
QY 2598 GCGCCCAATAGAGAGAGAGTCCAGTATCTCAGAGCGGTCTGGCGCCACACAGAGCT 2657
Db 55678 GCGCCGACCGCAAGAGAGAGTCCAGTATCTCAGAGCGGTCTGGCGCCACAAAGCT 55737
QY 2658 GATGAGATCTGAGTACCCACACTGCTGAGAGCTGATACGGAACATGGCGGTTCTGCTATGCT 2717
Db 55738 GATGAGGTGAGTACCCACCTGCTGAGCTGATACGGAACAGCGGTTCTGCTCATGGCC 55797
QY 2718 TGGGAGCAATCACAGGAGCATTGACACAACTTCGTTCAATTCGATGAGGTAAG 2777
Db 55798 TGGGAGCAAAACACAGACAGCCTTTCAGGAGCAACTACGTTCAATCTTCAGCTGAGGTAAG 55857
QY 2778 CGCCTCACTTCACTATATATATATATATCTCTCTGATTTCTCTT---TATTTGTTT 2834
Db 55858 CGCCTCACTTCACTATATATATATATATCTCTCTGATTTCTCTT---TATTTGTTT 55917
QY 2835 CAGTCAATGGATTTCTCTGAGTACTCTCTTCCAAAGTTTATTCACAGTGAAGTGTG 2894
Db 55918 CGTCAATGGATTTCTCTGAGTACTCTCTTCCAAAGTTTATTCACATAGTGAAGTGTG 55977
QY 2895 ATTTAAGTTTGGGAGGCTCAGAA--GTAATGTTGATGTTGATATATATTTTAACTG 2952
Db 55978 ATTTAAGTTTGGGAGGCTCAGGAAGTGTGTTGCAATGATATATATTTTGAAGTCTG 56037
QY 2953 CATTCATCTAAGGATAGAAACCAAAATAATTAATAATTCAGAAATGATTTTAC 3012
Db 56038 CATTCATCTAAGGATAGGAAACC-AAAAATTTGAAAAATTCAGAAATGATTTTAC 56096
QY 3013 --AAAAAAGAGTGTTCATGTAGTTGCAATTACATTTAGGATCAAGTCTAGAGTGTTCAT 3070
Db 56097 AAAAAAAGAGTGTTCATGTAGTTGCAATTACATGTAGGATCAGTCTAGAGTGTTCGT 56156
QY 3071 TTAGGATTTGATATGATAGGAGTAATGATGAGATAGCTTCTAGCA 3122
Db 56157 TTAGGATTTGATATGATAGGAGTAATGATGAGATAGCTTCTAGCA 56208

RESULT 3

AAF22305_09

Continuation (10 of 11) of AAF22305 from base 900001 (Arabidopsis thaliana chromosome
WP Sequence split into 11 fragments LOCUS AAF22305 Accession Aaf22305

WP	Fragment Name	Begin	End
WP	AAF22305_00	1	110000
WP	AAF22305_01	100001	210000
WP	AAF22305_02	200001	310000
WP	AAF22305_03	300001	410000
WP	AAF22305_04	400001	510000
WP	AAF22305_05	500001	610000
WP	AAF22305_06	600001	710000
WP	AAF22305_07	700001	810000

[illegible]

Db 26561 AGCGAGTTTGACAGGAGTGTCTCATTTTCAGTGAGCATGTACCTCCAGCGAGGAAAG 26620
Qy 2252 CAAGAGTTTGAGCGAAGCTACAGGAACAACAGCAAGCTGAGAGTGGTCCAGAAACA 2311
Db 26621 CAAGAGTTTGAGTGAAGCTACAGGAACAACAGCAAGCTGAGAGTGGTGAAGAAACA 26680
Qy 2312 GGATAAGTTACTCGCCAGTGCCTCAGGGCTATCAAGTTTCTGAAGACAAAGATCAGCTG 2371
Db 26681 TGATAAGTTACTCGCCAGTGCCTCAGGGCTATCAAGTTTCTGAAGACAAAGATCAGCTG 26740
Qy 2372 CTCCTCTTCCACTCAACTATTCGCAATGACAGCTCCCTCAGAGCATGCCCTTCAGAGAG 2431
Db 26741 CTCCTCTACCACTACAGCTATTCGCAATGACAGCTCCCTCAGAGCATGCCCTTCAGAGAG 26800
Qy 2432 ATATGACGC-----GCCGAGCTAGAGAGCAGAAATTCCTCA 2470
Db 26801 ATATGATCGCTGAGACAAGTCCGCGAGGCTCAGGCTACTGAGCAAGATATATCA 26860
Qy 2471 TGTCCCTCGAGGCAATTCATATTCGAGCTCGTGAATCTAGGAAGAATAGAGAACGAC 2530
Db 26861 TGTCCCTCGAGGCAATTCATATTCGAGCTCGTGAATTTGGAGAAAGAGGAGACCAA 26920
Qy 2531 ACTCACTCGATCTAGCAGAGAGCAGACGACTTCGAGTCTCTAGTTTACGCGACCG 2590
Db 26921 ACTTACTCGATCTAGCAGAGAGCAGACGACTTCGCAATCTCGTAGTTTACGCGACCG 26980
Qy 2591 CGGTCTGCCCAATAGAGAGAGAGTGCAGTATCTCAGAGCGGTGCTGGCGCCA 2650
Db 26981 CGG-----TAGAAGAGAGAGTGCAGTATCTCTCA-AGCGTGTGGCGCCA 27027
Qy 2651 CAGAGCTGATGATCGAGTACCCACATGCTCGAGCTGATACGGAACATGGCGGTTCGTC 2710
Db 27028 CAGAGCTGATGATCGAGTACCCACATGCTAGAGCTGATACAGAACAGGGCGGATCTTC 27087
Qy 2711 TATGGCTTGGAGCAATACAGGAGCGCAATGATGATGATGATGATGATGATGATGATG 2770
Db 27088 TATGGCTTGGAGCAATACAGGAGCGCAATGATGATGATGATGATGATGATGATGATG 27147
Qy 2771 AGGTAGCGCTCATTACCAATATATATATATATATATATATATATATATATATATAT 2830
Db 27148 AGGTAGCGCTCATTACCAATATATATATATATATATATATATATATATATATAT 27207
Qy 2831 GTTTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2890
Db 27208 GTTTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 27267
Qy 2891 TGTGATTTAAGTTTGGGAGGCTCAGGAAGTATGT 2927
Db 27268 TGTGATTTAAGTTTGGGAGGCTCAGGAAGTATGT 27304

RESULT 4
AAF22283
ID AAF22283 standard; DNA; 83390 BP.

AC AAF22283;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #6.

DE Centromere; microsome; vector; ds.

OS Arabidopsis thaliana.

XX WO20005325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000WO-US007392.

XX 18-MAR-1999; 99US-0125219P.

PR 01-APR-1999; 99US-0127409P.

PR 18-MAY-1999; 99US-0134770P.
PR 13-SEP-1999; 99US-0153584P.
PR 17-SEP-1999; 99US-0154603P.
PR 16-DEC-1999; 99US-0172493P.
XX (UYCH-) UNIV CHICAGO.
XX Preuss D, Copenhaver G, Keith K;
XX MPI; 2000-587529/55.
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for the
PT construction of transgenic plant and animal cells.
XX Claim 102; Page 386-404; 1449pp; English.
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for the
CC construction of transgenic plant and animal cells expressing selected
CC proteins such as hormones, enzymes, interleukins, clotting factors,
CC cytokines, antibodies, and growth factors
XX Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 0 U; 57 Other;
SQ

Query Match 50.9%; Score 1588.2; DB 3; Length 83390;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 2174; Conservative 0; Mismatches 263; Indels 358; Gaps 7;

Qy 662 CTTCTACATTCACACTCGACACCGGTGTCTACCATCTCCACACCGCTCTTTGTTCCAC 721
Db 38908 CTTCTTCAAGTCTACCATCTCGAC--CTACGATCTCTTACCAAGAACGACTTCACT 38965
Qy 722 ATACACTGACCAACAACCTCTCAAAAGTAAAAAAGAAAAAAGAAAAAAGTCAAAACCGAC 781
Db 38966 CGTCTGACGAATACCTTGTATCAACCGGAAGAAAAAAGAAAAAAGTCAAAACCGAC 39025
Qy 782 AGTTTCACTCAACCGGTTTCTCGACCGGTGCTGTTTGTAGATGTTGTTGTTGTTG 841
Db 39026 AGTTTCACTCGATTAATCACTCGACCGGTTTCTGTTGTTGTTGTTGTTGTTGTTG 39085
Qy 842 CTATTACTACATATTAAGTTTCTTCTGAGTTTCTGTTGTTTCTGTTTCTGTTTCTGTT 901
Db 39086 TTGAGACTAACTTATTAAGTTTCTTCTGAGTTTCTTCTGAGTTTCTTCTGAGTTTCT 39145
Qy 902 GTAACCTACAGTGGAAAAATCTCTATGGACCTGATTATATATGTTGATGAAGTAACTCT 961
Db 39146 GTAACCTACAGTGGAAAAATCTCTATGGACCTGATTATATATGTTGATGAAGTAACTCT 39205
Qy 962 GGTCCACTAGACCGGAGTGGAGCAACATGTTTACAGAGCTATAGGATGAATTTCAAC 1021
Db 39206 GGTCTCTAGACCGGAGAGAGAAACATGTTTACAGAGCTTTCAGGATGAATTTCAAC 39265
Qy 1022 GCTCTGACGTGACGTTATCAAGAGAGCTGAAATCGCTAGAGAAAGAGGCGCATGT 1081
Db 39266 GCTCTCTGCTCGAGTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 39325
Qy 1082 CGAGTAGATATGAGCTGATTGATGAGGATATCAAACTGAGTATGAGCCAGAGTCAATGGC 1141
Db 39326 CGAGTAGATATGAGCTGATTGATGAGGATATCAAACTGAGTATGAGCCAGAGTCAATGGC 39385
Qy 1142 GCAAG 1201
Db 39386 GCAAG 39445
Qy 1202 TCTTTGAGATGAATGATCTCTGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCAGT 1261
Db 39446 TCTTTGAGATGAATGATCTCTGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCAGT 39505
Qy 1262 TGGGTTTACTGGAGGAGCTGCAGCATCTGTTTCGAGAGTGTCTATCTG----- 1308
Db 39506 TGGGTTTACTGGAGGAGCTGCAGCATCTGTTTCGAGAGTGTCTATCTG----- 39565

QY 1309 -----ATAGGAGGAGACAAATCGAGTTCTTTCCACACTGCAAGTGG 1350
Db 39566 CTTACTCTTAACGCTTCCTATATAGGAGGAGACAAATCGAGTTCTTTCCACACTGCAAGTGG 39625
QY 1351 AAATGTATGAGGACTCACAGACTTTGAGCTGG-----1383
Db 39626 AGAATGATGAGGACTCACATACATTGAGTTTAGGGTTTAGGGTTTAGGGTTT 39685
QY 1384 -----1383
Db 39686 AGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGG 39745
QY 1384 -----1383
Db 39746 TTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTT 39805
QY 1384 -----1383
Db 39806 GGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGT 39865
QY 1384 -----1383
Db 39866 TTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAG 39925
QY 1384 -----ATACCATGGGTTAGGCTTCTTTGACGTTCTTTAGTGGAACAGCGGT 1431
Db 39926 GGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTT 39985
QY 1432 ACCAGATTAGATCAAGAAATTTGAAGAACTGTTTGTGTTTCCCTAGTGGAAAGGAACCA 1491
Db 39986 ATCAGCTTTCGATCAAGAGTTTGAAGAACTGTTTGTGTTTCCCTAGTGGAAAGGAACCA 40045
QY 1492 ACCCCAGTTTGAAGGAAAGAGCTTAAGGATTGTGGGCTACTATTGGGAAACAATCTAC 1551
Db 40046 AAACCAGTTTGAAGGAAAGAGCTTAAGGATTGTGGGCTACTATTGGGAAATTTTAC 40105
QY 1552 CGCTAACTCGACGGGTCAAGAGCAACCAAAATCCGAGTCTGTGATTCGCTACTTTTC 1611
Db 40106 CGCTAACTCGACGGGTCAAGAGCAACCAAAATCCGAGTCTGTGATTCGCTACTTTTC 40165
QY 1612 AGCGCTCGGTGTCMAATGTTTTTACTCCAGGAGTCTACAGGACCGGTCTCTAAACAG 1671
Db 40166 AGCGCTCGGTGTCMAATGTTTTTACTCCAGGAGTCTACAGGACCGGTCTCTAAACAG 40225
QY 1672 ACATGAGATGATAGATTCCGCGCTTACAGGATTTCTGCTGTACAAAGGGAAGATG 1731
Db 40226 ACATGAGATGATAGATTCCGCGCTTACAGGATTTCTGCTGTACAAAGGGAAGATG 40285
QY 1732 TCCAGAGAGAGATCTTAACGACTCACACAGTAAATGCTCTGTGATCCATCTGTGTG 1791
Db 40286 TCTTGAGAGGCGATCTCAACACGACCACTAGTCAATGCTCTGTGATCCATCTGTGTG 40345
QY 1792 GGTACATGAAGTGGCGCTGACAAACGCAAGGAAGGTAAGAGGACACTATGCGTGG 1851
Db 40346 GATAAGGAAGTGGCGGTGTACAAACGCACTAGTCAATGCTCTGTGATCCATCTGTGTG 40405
QY 1852 GTGGCGTTGTAGCCCAATCTGAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAG 1911
Db 40406 ATGGCGTTGTAGCCCAATCTGTGAAGTTGTGGAGTACCGCTTAGGAACCAAGGTTAG 40465
QY 1912 CACCGAAGATGAGACTTGGATCACTTGGCGGATGTGAGTTCCTGAGTTTGACATGG 1971
Db 40466 CACCGAAGATGAGACTTGGATCACTTGGCGGATGTGAGTTCCTGAGTTTGACATGG 40525
QY 1972 TTGGCGACTTTCACCGCTCAGGTTCCGAGCTTCAATGATAGATAGATTCGATCTTT 2031
Db 40526 CTAGCGACTTTCACCGCTCAGGTTCCGAGCTTCAATGATAGATAGATTCGATCTTT 40585
QY 2032 TCCCTGCAATTCAGCTACTAGGATTCTCGAGGAGGAGCAATGACTTCAAGCTGCGC 2091
Db 40586 TCCCTGCAATTCAGCTACTAGGATTCTCGAGGAGGAGCAATGACTTCAAGCTGCGC 40645

QY 2092 TTGAAGATCTTTATTTTCGAGGCGAGTCCGCAACTGAGAGATTAGTCAACCGAGGAG 2151
Db 40646 TTGAAGATCTTTATTTTCGAGGCGAGTCCGCAACTGAGAGATTAGTCAACCGAGGAG 40705
QY 2152 CTACAATAGAAAGATTTGATGAGACATATGATATAGATGAGCGGAGTTTGACACAGCA 2211
Db 40706 CTATAACAGAAAGATTTGATGAGACATATGATATAGATGAGCGGAGTTTGACACAGCA 40765
QY 2212 TGTATCATTTTCAGTGCAGATATACCTCCAGCGAGGAAAAGCAAGATTGAGCGAAGCTC 2271
Db 40766 TGTATCATTTTCAGTGCAGATATACCTCCAGCGAGGAAAAGCAAGATTGAGCGAAGCTC 40825
QY 2272 ACAGGAAACACAGCAAGCTGCAGAAAGTGTGCAAGAAACAGGATAAGTTTCTGCCAAGT 2331
Db 40826 ACAGGATAACATCAAGCTGCAGAAAGTGTGCAAGAAACAGGATAAGTTTCTGCCAAGT 40885
QY 2332 GCCTCAGGGCTATCAAGTTTCTGAAGGACAAAGATCAGCTGCTCTCTTCCACTACAACTA 2391
Db 40886 GCCTCAGGGCTATCAAGTTTCTGAAGGATAAGATCAGCTGCTCTCTTCCACTACAGCTA 40945
QY 2392 TTCCGCAATGACAGCTCCCTCAGGACATGCTTCGAGGAGATATGACGCGCC-----2444
Db 40946 TTCCACAGGACAGCTCCCTCAGGACATGCTTCGAGGAGATATGACGCGCCCGCGCA 41005
QY 2445 -----GAGCCTAGAGAGCAGAAAGATTCTGCAATGCTCCCTCGAGGAGCTTCAT 2490
Db 41006 TTCCGCGCAACCTGAGCCTAGAGAGCAGGAGATTCTGCAATGCTCCCTCGAGGAGCTTCAT 41065
QY 2491 CATTGAGGCTCGTGAATCTAGGAAGAAATAGGAACGACACTCCTCGATCTAGCAGCA 2550
Db 41066 CATTGAGGCTCGTGAATCTAGGAAGAAATAGGAACGACACTCCTCGATCTAGCAGCA 41125
QY 2551 GAGCAGACGCACTCTGCACTCTGCTAGTTTACGCGACCGGGTCTGGCGCGCAATGAA 2610
Db 41126 AGGTGAGCAGCTCTGCACTCTGCTAGTTTAC-----ATGAA 41163
QY 2611 GAAGAGAGTTCGAGTATCTCTCAGAGCGGTCTCGCGCCACAGAGCTGATGAGATCCAGT 2670
Db 41164 GAAGAGAGTTCGAGTATCTCTCAGAGCGGTCTCGCGCCACAGAGCTGAGAGTCCAGT 41223
QY 2671 ACCCACTACTGAGCTGATACGGAACAGCGCGGTCTGCTATGCTGCTGGGAGCAATCAC 2730
Db 41224 ACCCACTACTGAGCTGATACGGAACAGCGCGGTCTGCTATGCTGCTGGGAGCAATCAC 41283
QY 2731 AGCAGCCTATGAGTACCAACTCTGCTATTTTTCAGTGCAGTAAAGCGCTCAGTTCAC 2790
Db 41284 AGCAGCCTATGAGTACCAACTCTGCTATTTTTCAGTGCAGTAAAGCGCTCAGTTCAC 41343
QY 2791 CATTAATTATATCATCTCTGCTGATTTGTTCTTTATTTTTCAGTGCAGTAAAGTTGT 2850
Db 41344 CATTAATTATATCATCTCTGCTGATTTGTTCTTTATTTTTCAGTGCAGTAAAGTTGT 41403
QY 2851 CTGAGTACTCTCTTCCAAAGTTTATTCACAGTGGACTGTGTGATTTTAAAGTTTGGGGA 2910
Db 41404 CTGAGTACTCTCTTCCAAAGTTTATTCACAGTGGACTGTGTGATTTTAAAGTTTGGGGA 41463
QY 2911 GGGCTCAGGAAGTATG--TTGCAATGTATATTTTAACTGCTGCAATTCATCTAAGCAAT 2968
Db 41464 GGGCTCAGGAAGTATGATTTGCAATTTTATTTTAACTGCTGCAATTCATCTAAGCAAT 41523
QY 2969 AGAAAAACAAAAAATTTAAAAATTTTCAAGAAATGATTTTCCAC-AAAAAAGAGTGTTC 3027
Db 41524 AGAAAAACAAAAAATTTTAAAAATTTTCAAGAAATGATTTTCAAAAAAAGAAATGTTT 41583
QY 3028 ATGTAAGTGCATTAATTTAGGATCAAGTCTAGAGTGTTCATTTAGGATTTGTCATAT 3087
Db 41584 ATGTAAGTGCATTAATTTAGGATCAAGTCTAGAGTGTTCATTTAGGATTTTGCATAT 41643
QY 3088 GCATAGGAGTAATGATGAGATAGCTTGTAAAGCA 3122
Db 41644 GCATAGGAGTAATGATGAGATAGCTTGTAAAGCA 41678

Db 53040 TCCTGAGAGCGATCTCAACAGCGCACCTAGTATGCCCTCTGTGTATCCACTGTGTG 53099
Qy 1792 GGTTCATGAAGTGGCGCTGCAAAACGCAAGAAAGGTAAAGAGGACACTATGCGTGG 1851
Db 53100 GATATAGGAAGTGGCGCTGTACGAACAGCAAGAAATAGGTGACAGGACACTATGCAATG 53159
Qy 1852 GTGGCGTGTGACGCCAATCTGAAAGTTTGTGAGTTCCCTCAAGGAAGTAGGTTAG 1911
Db 53160 ATGGCGTTGTGACGCCAATCTGAAAGCTTGTGAGTACCGCTTAAGGAACACGAGTTAG 53219
Qy 1912 CACCGAGAAATGATGGACTTGGATCACCTTGCCTGATGTGAGTTCTCTGAGTTTGCACATGG 1971
Db 53220 CACCGAGAAATGATGGACTTGGATCACCTTGCCTGATGTGAGTTCTCTGAGTTTGCACATGG 53279
Qy 1972 TTGGCGACTTTCACCGCTACAGTTTCGAGCAATTCATCGAATPAGAAATCGCCAACTTCTTT 2031
Db 53280 CTAGCGACTTTCACCGCTACAGTTTCGAGCAATTCATCGAATPAGAAATCGCCAACTTCTTT 53339
Qy 2032 TCCCTCGAATTCAGCTACTAGGATCTCGAGGCGAGAACTTGACTTCAAGCCTGCGC 2091
Db 53340 TCCCTCGAATTCAGCTACTAGGATCTCGAGGCGAGAACTTGACTTCAAGCCTGCGC 53399
Qy 2092 TTGAAGATCTTTATTTCCAGGCGAGTCCGCCAATCTGAGGAGATTAGTCACACCGAAGGAG 2151
Db 53400 TTGAAGATCTTTATTTCCAGGCGAGTCTGCCAATCTGAGGAGATTAGTCACACCGAAGGAG 53459
Qy 2152 CTACAAATGAAGATGTTGATGAGACATATATATATAGATAGGCGGAGTTTGACACGAGCA 2211
Db 53460 CTATAACAGAAAGATGTTGATGAGACATATATATATAGATAGGCGGAGTTTGACACGAGCA 53519
Qy 2212 TGTATCATTTTCAGTGAGCATATACCTCCAGGAGGAAAGCAAGAGTTTGAGCGAAGCTC 2271
Db 53520 TGTATCATTTTCAGTGAGCATATACCTCCAGGAGGAAAGCAAGAGTTTGAGCGAAGCTC 53579
Qy 2272 ACAGGAACACAGCAAGCTCGAAGATGGTGCAGAAACAGGATAAGTTACTCGCCCAAGT 2331
Db 53580 ACAGGAATACATCAAGCTCGAAGATGGTGCAGAAACAGGATAAGTTACTCGCCCAAGT 53639
Qy 2332 GCCTCAGGGCTATCAAGTTTCTGAAGCAAGATCAGCTCTCTCTTCCACTACACTA 2391
Db 53640 GCCTCAGGGCTATCAAGTTTCTGAAGCAAGATCAGCTCTCTCTTCCACTACACTA 53699
Qy 2392 TTCGGAATCAGAGCTCCCTCAGGACATGCTTCGAGGAGATATGAGCGCCCTC 2444
Db 53700 TTCACAGGACAGCTCCCTCAGGACATGCTTCGAGGAGATATGAGCGCCCTCAGGACAA 53759
Qy 2445 -----GAGCCTAGAGAGCAGAAAGATTCTGCAATGTCCTCGCAGGCAATTCAT 2490
Db 53760 TTCGCGCAACCTGAGCCTAGAGAGCAGAGATTCTGCAATGTCCTCGCAGGCAATTCAT 53819
Qy 2491 CATTGAGCCTCGTGAATCTAGGAAGATAGGAGAACGACACTCACTCGATCTAGCAGCA 2550
Db 53820 CATTGAGCCTCGTGAATCTAGGAAGATAGGAGAACGACACTCACTCGATCTAGCAGCA 53879
Qy 2551 GGAGCAGACGACTTCTGCACTCTGAGTTTACGCGACCGGCTGCGCGCAATAGAA 2610
Db 53880 AGGTGACACACTCTGCACTCTGAGTTTACGCGACCGGCTGCGCGCAATAGAA 53917
Qy 2611 GAAGAGAGTCTGAGTATCTCTCAGAGCGGCTCGCGCGCACAGAGCTGATGAGATCGAGT 2670
Db 53918 GAAGAGAGTCTGAGTATCTCTCAGAGCGGCTCGCGCGCACAGAGCTGATGAGATCGAGT 53977
Qy 2671 ACCCATGCTGAGCTGATACGGAACATGCGGTTTCTGTATGCTTGGGAGCAATCAC 2730
Db 53978 ACCCATGCTGAGCTGATACGGAACATGCGGTTTCTGTATGCTTGGGAGCAATCAC 54037
Qy 2731 AGCAGCCATGACTACCACTTCTGTTTATTCGACTGAGTAAAGCGCTCACTTCAC 2790
Db 54038 AGCAGCCATGACTACCACTTCTGTTTATTCGACTGAGTAAAGCGCTCACTTCAC 54097
Qy 2791 CATTATATATATCATCTCTGTGATTTGTTTCTTTTATTTTCTGATGATTTGTT 2850
Db 54098 CATTGATATATATCATCTCTGTGATTTGTTTCTTTTATTTTCTGATGATTTGTT 54157

Qy 2851 CCTCAGTACTCTCTTCAAGTTTATTCACAGTGGACTGTGTGATTTAAGTTTGGGGA 2910
Db 54158 CCTGGTACTCTCTTCAAGTTTATTCACAGTGGACTGTGTGATTTAAGTTTGGGGA 54217
Qy 2911 GGGCTCAGGAAGTATG--TTGCATTGTATATATTTTAAAGTCTGCATTCACTCAAGCAT 2968
Db 54218 GGGCTCAGGAAGTGTATTTGCTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 54277
Qy 2969 AGAAAAACCAAAAAAATTTAAAAATTTTCAAAAAATGATTTTCAAAAAAAGATGTTT 3027
Db 54278 AGAAAAACCAAAAAAATTTTAAAAATTTTCAAAAAATGATTTTCAAAAAAAGATGTTT 54337
Qy 3028 ATGTAGTTGCTATTTACATTTTAGGATCAAGTCTAGAGTGTTCATTATTAGGATTTGTCATAT 3087
Db 54338 ATGTAGTTGCTATTTACATTTTAGGATCAAGTCTAGAGTGTTCATTATTAGGATTTTGCATAT 54397
Qy 3088 GCATAGGGGATATATGATGAGATAGCCTTTGTAAGCA 3122
Db 54398 GCATAGGGGATATATGATGAGATAGCCTTTGTAAGCA 54432

RESULT 6

AAF22296/c
ID AAF22296 standard; DNA; 129021 BP.

XX AAF22296;

XX AC AC (first entry)

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #19.

XX Centromere; chromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO20005325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US007392.

XX 18-MAR-1999; 99US-0125219P.

XX 01-APR-1999; 99US-0127409P.

XX 18-MAY-1999; 99US-0134770P.

XX 13-SEP-1999; 99US-0153584P.

XX 17-SEP-1999; 99US-0154603P.

XX 16-DEC-1999; 99US-0172493P.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhagen G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for producing stably inherited chromosomes which can serve as vectors for the construction of transgenic plant and animal cells.

XX Claim 102; Page 686-716; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant (Arabidopsis thaliana) centromere. The constructs are useful for producing stably inherited chromosomes which can serve as vectors for the construction of transgenic plant and animal cells expressing selected proteins such as hormones, enzymes, interleukins, clotting factors, cytokines, antibodies, and growth factors

XX Query Match 45.3%; Score 1415.6; DB 3; Length 129021; Best Local Similarity 86.3%; Pred. No. 0; Matches 1610; Conservative 0; Mismatches 204; Indels 52; Gaps 2;

[illegible]

PD 22-MAR-2001.
 XX
 PF 14-SEP-2000; 2000WO-EP008994.
 XX
 PR 16-SEP-1999; 99GB-00021964.
 XX
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 PA
 XX Steimer A, Mittelsten Scheid O, Paszkowski J;
 PI WPI; 2001-244815/25.
 XX
 DR Novel transcriptionally silenced genes obtained from Arabidopsis useful
 XX for selecting plants which compared to wild type plant are impaired in
 PT transcriptional gene silencing.
 PT
 XX
 PS Claim 1; Page 26-28; 36pp; English.
 XX
 CC This present sequence is that of a 'virtual' cDNA combining overlapping
 CC sequences from 5' and 3' extensions (see AAF30450-53) of an Arabidopsis
 CC thaliana transcriptionally silent information (TSI) sequence, i.e. TSI-A
 CC (see AAF30454). The 4860 bp sequence was used in Arabidopsis genomic
 CC database searches, identifying a genomic sequence (see AAF30459) that
 CC included a transcribed region. Comparison of transcriptional gene
 CC expression between an Arabidopsis line carrying a silent transgene
 CC present in multiple copies and its mutant derivative moml impaired in
 CC silencing of the transgene revealed 2 cDNA clones which were expressed in
 CC the mutant plants, but not in the parental or wild-type plants. Both
 CC clones were derived from the same family of transcripts referred to as
 CC TSI. Genomic templates encoding TSI are repetitive elements with mainly
 CC pericentromeric location and conserved organisation among various
 CC ecotypes. Transcriptional silencing of genomic TSI templates is
 CC specifically released in the mutant. Transcription of TSI can be used as
 CC a marker to identify a defective silencing pathway in a plant. The
 CC invention provides a nucleic acid encoding genetic information which is
 CC silenced in wild-type plants but whose expression is turned on in plants
 CC defective in transcriptional gene silencing, and has the formula: RA-RB-
 CC RC, where RA, RB, RC constitute component nucleotide sequences selected
 CC from the present sequence and those given in AAF30450-59. Also provided
 CC are methods of selecting a plant that is impaired in transcriptional gene
 CC silencing, and of producing DNA representing at least part of a gene
 CC necessary to maintain silencing of another gene in a cell or plant
 XX
 SQ Sequence 4860 BP; 1296 A; 1068 C; 1131 G; 1365 T; 0 U; 0 Other;
 Query Match 44.8%; Score 1398.6; DB 4; Length 4860;
 Best Local Similarity 76.2%; Pred. No. 0;
 Matches 1836; Conservative 0; Mismatches 489; Indels 84; Gaps 6;
 QY 788 ACTCAACCGGTTTACTCGACCGGTACGCTGGTTTAGATTGTTTGGTTTGGTTATTA 847
 DB 328 ACTCGACCGGTTTCACTCGATCGGTTACTTGACT--GGTTAGTGTGTGTTTATTGAA 385
 QY 848 CTAACATATTAGCTTTATCTTTAGTTTC-GTCTGTTTTAGTTTCATCATGATAC 906
 DB 386 CTAACATATTGATTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 445
 QY 907 TACAGTGGAAATPCCTCTATGGACCTGATTAATGATGATGAAGTAACTCCTGCTC 966
 DB 446 TACAGTGGCAATCCTCCATGGATCGGATTACAGCTGATGAAGTGAATCTTGGTCA 505
 QY 967 ACTAGCCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGAGTAATTTGAACGCTCT 1026
 DB 506 ACTAGACCAGAGAGAGCAACAGGCTTATGAGAGCTTCAGAGCCGAGACCCCAAGCTCA 565
 QY 1027 GCAGCTCGAGCTAATCAACAGAGCTGAATCGCTAGAGAAAGAGGGCGATGTCAGT 1086
 DB 566 GTAGCTCGAGCAATGAAGAGAGCTGAGATTGCTAGAGAAAGAGAGCGATGACCA 625
 QY 1087 AGATATGAGCTGATTGATGAGGATATCAAACTGATGATGAGCCAGGTCATGGCGCAAG 1146
 DB 626 AGATATGAGTTGATCGACCAAGATTTGAGCTGAGTATGAGCCTGAGTCATGGCCACA 685

QY 1147 GAGACGAAGCTTACTGAACAAATCCGACGAGTTTACAGTGGAGGAGTATATCAGATTCTTT 1206
 DB 686 GAAACAAAATCTGTTGAACACAGCCTGATGAAGTTTACAGTGGAGAGATCATCAGATTCTTC 745
 QY 1207 GAGATGAATGACTTCTGGGAAACGAGGTATCCCTGATATGAGACTTTTACGCCAGTTGGGG 1266
 DB 746 GAGCTGAACGACTTCTGGGAGCGAGGTACCCCTGTTATGAGACTCTAGAGCCAGCTTAGG 805
 QY 1267 TTACTGGAGGACGTCGAGCATCTGTTTCGAGAGTGTCTAT----- 1305
 DB 806 CTACTGGAGGACGTCACAGCACTTATTCGAGAGTGCATCTTTGAGAGCCTGATGCTTTAC 865
 QY 1306 -----CTGATAAGGAGGAGACAATCAGATTCTTTCCACACTGCAAGTGGAAATG 1355
 DB 866 CCGTACGTCGCTTACAAGAGGAAACATAGAGTTTCTCCACTCTGCAAGTGGAGTTG 925
 QY 1356 TATGAGGAGCTCACAGACTTTGAGCTGGATACCATGGGTTAGGCTTCTTGTGACCTTTCTTA 1415
 DB 926 TATCAGGAGCTTACTGCGAGATGAACCTGAGAGTGAAGGTTGGGATTTCTTGACTTTTCA 985
 QY 1416 GTGGATGAACAGCGGTACCAAGATTAGATCAAGAAATTTGGAAGACTGTTTGGTTTCCCT 1475
 DB 986 GTTAACAGCAGCGGTTACCAAGCTATCTATCAAGAGCTTGAAGAGATTATTTGGTTTCCC 1045
 QY 1476 AGTGGAAAGGAAACCAACCCAGGTTTGACAGGGAAGAGCTTAAAGGATTTGTGGGCTACT 1535
 DB 1046 AGTGGAAAGGAACTAAACCAAGTTGCAAGGGAAGAGTTGAAGATTTGTGGTTAAC 1105
 QY 1536 ATTGGGAAACAACTTACCGCTAAACTCGACCGGTTCCAAAGAGCAACCAAAATCCCGAGTCT 1595
 DB 1106 ATTGGGAAACGATTGGCGCTCAACTCTGCAAGGCTTAAGAGCAACAGATTTCGAAGCCCT 1165
 QY 1596 GTGATTGCTACTTTTTCAGCGCTCGGTTGCCAATGTTTTTACTCCAGGAGTCTACAGGC 1655
 DB 1166 GTGATCCGCTACTATCAGCGCTCAGTAGCGAATGTTCTGTACCCCAAGGAAATCTACAGGC 1225
 QY 1656 ACCGTGTCTAACACAGACATGAAGATGATAGTTTACGCGCTTATAGGATTTCTCCGCTT 1715
 DB 1226 ACCGTGTCTAACACAGACATGAAGATGATGATTTGCTGACTCAAGGTTATTTCCCGAGA 1285
 QY 1716 ACAAAGGAAAGAAATGCTCTGAGAGGAGATCTTAAAGACTTCAACCAAGTAAATGCTCTG 1775
 DB 1286 ACAAAGGAAAGAAAGTCTCTTAAAGGCGACCTTAAATGATACACACCGGTCATGCTCTG 1345
 QY 1776 TTGATCCATCTGTGCTGATCATGAAGTGGCGCTGACAAACGCGAAGAGAGGTAAGA 1835
 DB 1346 TTGATCCATCTGTGCTGATCATGAAGTGGCGCACACCAACGCGGAGGAGAGAGTGCGA 1405
 QY 1836 GGAGCACTATGCGTGGGTCGCTGTGTGACGCGCAATTTCTGAAAGTTTTGTGGAGTTCGCTC 1895
 DB 1406 GGAGCCCTTTGTGCTGGTGGCTGTGTGACACCGATTTCTGATTCATGTTGTGTTACCTCTC 1465
 QY 1896 AGGAAGTGGGTTAGCACCGAGATGATGGACTTGGATCACTTGGCCCGATGTGAGTTC 1955
 DB 1466 ACCTCTCCAGGTTTATCGAGGATGATGGATTTAGATCATTGCTGCTGTTGAGTTT 1525
 QY 1956 TCTGAGTTTGCATGTTGGCGACTTTTACCGCTACAGGTTTCAGGATTCATGATTTAG 2015
 DB 1526 CTGGAGTACGATGTTGGCGATTTCTATCGCTACAAATTTTCGAGACTCCCTGACCCGA 1585
 QY 2016 ATCGCCAAACATTTCTTTTCCCTGCAATTTTACGCTACTAGGATTTCTCGAGGCGAGAACAT 2075
 DB 1586 ACAGCCAAACATTTGCTTCCCTGCAATTCGAGGCCACAAACCTACTTTTAGGGTGAGAACAT 1645
 QY 2076 GACTTCAAGCCTGCGCTTGAAGATCTTTATTTTCGAGGCGAGTCCGCAACTCAGGAGATT 2135
 DB 1646 GACTTCAAGCCTGCGCTGATTTACCTCTTACTTTGAGAGCACTCCACCGACTCATGACAT 1705
 QY 2136 AGTCAACCGAGGAGCTACAAATAGAGATGTTGATGAGACATATGATATAGAGCG 2195
 DB 1706 GTCCCTACGAGGAAGCTACAGAGGATGATTTTGTGAGCGGATGAGATAGGAGGAG 1765
 QY 2196 GAGTTTGACACGAGCATGTATCATTTTTCAGTGAGCATATACCTCCAGCGGAGGAAAGCAAG 2255

Db 19158 GAGACGAAGCTGTTGAA CAGGCTTAATGAAGTCCAGTGGAGGAGTACATCAGACTTTTC 19217
 QY 1207 GAGATGAATAGTCTCTGGGGAACGAGGTATCCCTGATATGAGACTTTAGCCAGTTGGG 1266
 Db 19218 GAGCTGAACGACTCTCTGGGGAACGAGGTATCCCTGATATGAGACTTTAGCCAGTTGGG 19277
 QY 1267 TTACTGGAGGACGTCAGCATCTGTTCCAGAGTGTCTATG----- 1308
 Db 19278 CTACCGGAGGACGTACAA CACTTATTCGAGAAGTGTCTATCTGGAGACACTGATGTCTTAC 19337
 QY 1309 -----ATAAGGAGGAGACAAATCAGTCTGATACCAATGGGTTAGGCTTCTTACGCTTCTTA 1415
 Db 19338 CCGTACGTCGCTTACAGAAGGAATAAGAGATTTCTCCACTCTCTCAAGTGGAGATG 19397
 QY 1356 TATGAGGACTCAGACACTTTGAGCTGGATACCAATGGGTTAGGCTTCTTACGCTTCTTA 1415
 Db 19398 TATCAGGACTTACAGCAGATGAGCTGGAGATGAAGGTTGGGTTCTTGACTTTTTC 19457
 QY 1416 GTGATGAACAGCGGTACCAATTTAGATCAAGAAATTTGGAAGAACTGTTTGGTTCCCT 1475
 Db 19458 GTGAGCGAGCAGTGTATACAGTATCTATCAGGAGCTTGAAGGATTTGTTGGCTTACC 19517
 QY 1476 AGTGGAAAGGAACCAACCCAGCTTTGACAGGGAAGCTTGAAGGATTTGTTGGCTTACC 19517
 Db 19518 AGTGGAAAGGAACCAACCCAGCTTTGACAGGGAAGCTTGAAGGATTTGTTGGCTTACC 19517
 QY 1536 ATTGGGAACAACTTACCGCTAACTCGACGCGTCCAAAGAGCAACCAAAATCCGAGTCCCT 1595
 Db 19578 ATTGGGAACCAATGCGCTCACTCTCGAGGTCCAAAGAGCAACCAAAATCCGAGTCCCT 19637
 QY 1596 GTGATTCGCTACTTTCAGCGCTCGGTTGCCAATGTTTTTACTCCAGGAGTCTACAGGC 1655
 Db 19638 GTGATTCGCTACTTTCAGCGCTCGGTTGCCAATGTTTTTACTCCAGGAGTCTACAGGC 1655
 QY 1656 ACCGTGTCTAACACAGACATGAAGATGATAGTTACAGCGCTTATAGGATTTCTCCGCTT 1715
 Db 19698 ACCGTGTCTAACACAGACATGAAGATGATAGTTACAGCGCTTATAGGATTTCTCCGCTT 1715
 QY 1716 ACAAAGGAAAGATGCTCTGAGAGGAGATCTTAAACGACTCACACCAAGTAATGCCTCTG 1775
 Db 19758 ACAAAGGAAAGATGCTCTGAGAGGAGATCTTAAACGACTCACACCAAGTAATGCCTCTG 19817
 QY 1776 TTGATCCATCTGTGGGTACATGAAGTGGCGCTGACAAACGCGCAAGAAAGGTAAAG 1835
 Db 19818 TTGATCCATCTGTGGGTACATGAAGTGGCGCTGACAAACGCGCAAGAAAGGTAAAG 1835
 QY 1836 GGAGCACTATGCGTGGGTGGCTGTGACGCCAATCTGAAAGTGTGAGAGTTCGCTC 1895
 Db 19878 GGAGCCCTTTGTGTAGTGTGTGTGACACCAATCTGATTAACATGTGTGTGATCTTTC 19937
 QY 1896 AAGGAAGTAGGTTAGCACCGAAGTATGAGACTTTGATCACTTGGCCGATGTGAGTTC 1955
 Db 19938 ACCTCTCCAGGTTGATCCGAGGATGATGATTTAGATCACTTGGCTGTGTGATTT 19997
 QY 1956 TCTGAGTTGACATGTTGGCGACTTACCGCTACAGTTTCAGGATTCATCGATTA 2015
 Db 19998 CTGAGACACACATGTTGGCGATTTCTATCGCTACAAATTTGAGCACTCTTGGACCGGA 20057
 QY 2016 ATCGCCAAATCTTTTCCCTCGCATTTACGCTACTAGGATTTCCGAGGCGAGACAT 2075
 Db 20058 ACAGCCAAATCTTTTCCCTCGCATTTCTATCGCTACAAATTTGAGCACTCTTGGACCGGA 20117
 QY 2076 GACTTCAAGCTCGCTTGAAGATCTTTTATTTTCGAGGCGAGTCCGCCAATCAGGAGAT 2135
 Db 20118 GACTTCAAGCTCGCTGATTAACCTCTACTTTGAGGCGCTCCACGAGTATGACA-A 20176
 QY 2136 AGTCACCGAGGAGCTCAATAGAGATGTTGATGAGACATATGATATAGATGAGCG 2195
 Db 20177 CGTCCCTCAAGAGAGCTACCAAGCTGAGATTTGCTGAGACATGAGGATAGGAGGAG 20236
 QY 2196 GAGTTTGAACAGCAGATGATATCTTTTCACTGAGCATATACCTCCAGGAGGAAAGCAAG 2255
 Db 20237 GAGTACGATACGCGCATGATATCTTTTCACTGAGCATGTTACTCCAGGAGGAGCAAG 20296

QY 2256 AGTTTGACGGAAGCTCACAGGAACAA CAGAGCTGCAAGACTGGTCAAGAAACAGGAT 2315
 Db 20297 AGCTTGACGGAAGCTCACAGGAACAA CAGAGTGAAGTTGCAAGAGTGGTCAAGAAACAGAT 20356
 QY 2316 AAGTTACTCGCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAAAGTCAAGCTGCTTCC 2375
 Db 20357 AGGTACTCATCAAGTGTCTCAAGCCATCAAGTTTCTGACAGACAAAGTCAAGCTGCTTCC 20416
 QY 2376 TCTTCCACTCAACTATTTCCCAATGACAGCTCCCTCAGGACATGCTTCCAGGAGATAT 2435
 Db 20417 TCTTCTACACAGGATTTCCGACGAGAGCTCTCTCAGGACATGCTTCCAGGAGATAT 20476
 QY 2436 GAGCC-----GAGCC-----GCCCGAGCCCTAGA 2453
 Db 20477 GAGCGCCAGAGCCAACTCGCCACAGGCTGAGCCAAAGTCAACACATGCTTGAAGCTAGT 20536
 QY 2454 GAGCAGAAGATTTCTGATGCTCCCTGCGAGGCTTTCATCATTCGAGCTCGTGAATCTAGG 2513
 Db 20537 GAGCGAGTAGTCTCACAAGTCCCTGCGTCTGGCAGCAGGAGTACACGACTTCTCCAGTCC 20596
 QY 2514 AAGNATAGGAGAACGACACTCACTCGATCTAGCAGGAGCAGACGACTTCTCCAGTCT 2573
 Db 20597 AGAAAGAAAGAGGCTGCACTCGCTCTGGCAGCAGGAGTACACGACTTCTCCAGTCC 20656
 QY 2574 CGTAGTTTACCGACCGCGGTGCTGCGGCAATAGAAAGAGAGAGGTCGATATCTCTCAG 2633
 Db 20657 CGTAGTTTACCGATCGCGGTGCTGCGGCAATAGAAAGAGAGATCGAGTATCATCTCAG 20716
 QY 2634 AGCGGTCTGCGCCCAACAGAGCTGATGAGATCGATACCCACATGCTGAGCTGTATACG 2693
 Db 20717 AGCGGTCTGCGCCGAGTAAAGAGCAGAGGTTGAGTACCCAGCGGAGCTGAGACA 20776
 QY 2694 GAACATGCGGTTGCTCTATGCTTGGAGCAATCACAGGAGCCATGTACTACCAACTT 2753
 Db 20777 CAAACAGAGATTTCTCGATGGCTTGGAGCAITTCACAGCGGCTGTTGATGACCAACTC 20836
 QY 2754 CGTTTCAATTTCCAGTGAAGTGAAGCGCTCACTTCCACTATATATATATCATCTCTGT 2813
 Db 20837 CGTCTCTTCTCCAGTGAAGTGAAGCGCTCACTTCCACTATATATATATATCATCTCTGT 2813
 QY 2814 GATTTGTTCTTTATTTTTCAGTGATTTGATTTGCTCTGAGTACTCTCTTCCAACTTT 2873
 Db 20896 -----TTTATTTTGTCTGTGATGTTTGTCTGAGTACTCTCTTCCAAATTT 20948
 QY 2874 ATTACACAGTGAAGTGTGATTTAAGTTTGGGAGGCTCAGGAA--GTATGTTGCA 2931
 Db 20949 GGTACACAGTGAAGTGTGATTTAAGTTTGGGAGGCTCAGGAAAGTGTGTTGCG 21008
 QY 2932 TTGTATATATTTTAAAGTCTGCAATTCATCTAAGGCATAGAAAACC---AAAAAAATTT 2998
 Db 21009 TTGTATATATCTTGAAGTCTGCAATTCATCTAAGGCATAGAAAACC---AAAAAAATTT 21068
 QY 2989 AAAAAATTCAGAAAATGATTTCAAAAAAAGAGTGTTCATGTAGTTGCAATACATTTAG 3048
 Db 21069 AAAAAATTCAGAAAATGATTTCAAAAAACAGAGTGTTCATGTAGTTGCAATTTAG 21128
 QY 3049 GATCAAGTCTAGAGTGTTCATTTAGGATTTGCAATGCAATGAGGAGATAATGATGAGA 3108
 Db 21129 GATCGAGTCTAGAGTGTTCATTTAGGATTCATTTGATATGATGAGGAGCAATGATGAGA 21188
 QY 3109 TAGCCTTTGAAGCA 3122
 Db 21189 TAGCCTTTGAAGCA 21202

RESULT 9
 AAF22279
 ID AAF22279 standard; DNA; 64415 BP.
 XX AAF22279;
 XX
 DT 20-MAR-2001 (first entry)

Db	3487	AGCTTGAGCGAAGGCTCAGAGAAACAACAGTAAGTTGCAGAGGTGTCGCAAGAACAGAT	3546
QY	2316	AAGTTACTCGCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGGACAGATCAGCTGCTCC	2375
Db	3547	AGGCTACTCATCAAGTGCTTCAAGGCCATCAAGTTTCTGACAGACAAGCTAAGCTGCTCC	3606
QY	2376	TCTTCCACTACAACTATTTCGCAATGACAGCTCCCTCAGGACATGCCTTCGAGGAGATAT	2435
Db	3607	TCTTCTACCAAGCTATTCCGAGGGAGAGCCTCCTCAGGACATGCCTTCGAGAGATAT	3666
QY	2436	GACGC-----GACGC-----GCCCGAGCCTTAGA	2453
Db	3667	GACGCGCAGAGCCAACTCGCCACTGGCCTGAGCTAAGTCACACAGGCCTGAGCCTAGT	3726
QY	2454	GAGCAGAGAGATTCTGCATGTCCCTCGAGGGAATTCATCATTCGAGCCTCGTGAATCTAGG	2513
Db	3727	GACCGAGTAGTCCCACAGCTCCCTGTGTGGCAITTCATCAITCAAGCCTCGGAGCTCGGG	3786
QY	2514	AAGAATAGGAGAACGACACTCACTCGATCTAGCAGCAGGAGCAGACGACTTCGCGAGTCT	2573
Db	3787	AGAAAGAGAGAGGCTGCACCTCGCTCGCTGTGGCAGTAGGAGTACACGACTTCTCCAGTCC	3846
QY	2574	CGTAGTTTACGCGACCCCGGTGCTGGCCGCAATAGAAAGAGAGAGTTCGAGTATCCTCAG	2633
Db	3847	CGTAGCTTACGCAACACCGGTGATGGCCGCGACGATAGAAACAGAGGTTCGAGTATCATCAG	3906
QY	2634	AGCGTGCTGGCCGCCACAGAGCTGATGAGATCGAGTACCACATGCTGAGCTGATACG	2693
Db	3907	AGCGCTCTGGCCGCGACGAAGGAGCAGAGGTTCGAGTACCCTCCAGGGGAGCTAAGACA	3966
QY	2694	GACATGGCGGTTGCTCTATGGCTTGGGAGCAATCACAGGCAGCCATTCGACTACCAACTT	2753
Db	3967	CACAGGGAGATTCTTCGATGGCCTGGGAGCAATCACAGCAGCTATTGACGACCACTC	4026
QY	2754	CGTTCTATTTCGACTGAGGTAAGCGCCTCACCTTCACCATATATATATATCATCTCTGT	2813
Db	4027	CGCTCCTCTTCCACTGAGGTAAAGCACTATCTCCACCATTTGAATATACCATCTCCTG-	4085
QY	2814	GAATTTGTTCTTTATTTTGTTCAGTGATTTGGATTTGTCCTGAGTACTCTCTTCCAAAGTTT	2873
Db	4086	-----TTTTTATTTTGTGATGTGTTTGTCTGAGTACTCTCTTCCAAATTT	4138
QY	2874	ATTTCACAGCTGACTGTGTGATTTAAGTTTGGGGAGGGCTCAGAA--GTATGTGCA	2931
Db	4139	GGTCACACAGTGTGACTGTGTGATTTAAGTTTGGGGAGGGCTCAGGAAGTGTGTGCA	4198
QY	2932	TTGTATATATTTTAACTCTGCATTCATCTAAGGCATAGAAAAACAAAAAAATTAA	2991
Db	4199	TTGTATATATCTTGAGTCTGCATTCATCTGAGCATAGAAAAACAAAAAAATTGAA	4258
QY	2992	AATTTCAGAAAAATGATTTACAAAAAAGAGTCTTCATCTGATGTGATTCACATTTAGGT	3051
Db	4259	AATTTCAGAACATGATTTACAAAAATAGAGTGTTCATGTGTGATTCGATTTAGGAT	4318
QY	3052	CAAGTCTAGAGTGTTCATTTAGGATTTGTTCATATGCATAGGGGATATGATGAGATAG	3111
Db	4319	CGAGTCTAGAGTGTTCGTTTAGGATTTGTTCATATGCATAGGGGATATGATGAGATAG	4378
QY	3112	CCTTGTGAAGCA 3122	
Db	4379	TCTTGTGAAGCA 4389	

RESULT 10
AAF22283/C
ID AAF22283 standard; DNA; 83390 BP.

AA
AC AAF22283;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #6.

XX	Centromere; microsome; vector; ds.	
KW	Arabidopsis thaliana.	
OS	WO200055325-A2.	
PN	21-SEP-2000.	
XX	17-MAR-2000; 2000WO-US007392.	
PF	18-MAR-1999; 99US-0125219P.	
XX	01-APR-1999; 99US-0127409P.	
PR	13-MAY-1999; 99US-0134770P.	
PR	18-SEP-1999; 99US-0153584P.	
PR	17-SEP-1999; 99US-0154603P.	
PR	16-DEC-1999; 99US-0172493P.	
XX	(UYCH-) UNIV CHICAGO.	
PA	Preuss D, Copenhaver G, Keith K;	
XX	WPI; 2000-587529/55.	
DR	Recombinant DNA construct comprising a plant centromere, useful for	
PT	producing stably inherited microsome which can serve as vectors for the	
PT	construction of transgenic plant and animal cells.	
XX	Claim 102; Page 386-404; 1449pp; English.	
PS	The present invention relates to a recombinant DNA construct of a plant	
XX	(Arabidopsis thaliana) centromere. The constructs are useful for	
CC	producing stably inherited microsome which can serve as vectors for the	
CC	construction of transgenic plant and animal cells expressing selected	
CC	proteins such as hormones, enzymes, interleukins, clotting factors,	
CC	cytokines, antibodies, and growth factors	
XX	Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 0 U; 57 Other;	
SQ	Query Match 42.0%; Score 1311; DB 3; Length 83390;	
	Best Local Similarity 75.2%; Pred. No. 0;	
	Matches 1809; Conservative 0; Mismatches 470; Indels 126; Gaps 9;	
QY	795 CGGTTTACTCGACCGGTACGCTGGTTAGATTGTGTTTGGTTTGGTTTGGTTTACTTAACAT 854	
Db	23635 CCGGTTACTCGACCGGTAACTTGATT--GGTTTAGTTGTTGGTTTATTGGACTTAACAT 23578	
QY	855 ATTAAAGTTTAUCTTTGAGTTTCGTCGTTTTTAGG-TTTCATCATGAGTAACACTACAGTG 913	
Db	23577 ATTGATATTGGTTTAAAGTTACATTATTTTCAGAAATCAATATGATGATTAACAGTG 23518	
QY	914 GAAATTCCTTCTTGGACCTGATTATATGTCGATGAAGCTAAGTCCTGGTCCATAGAC 973	
Db	23517 GCGATTCCTCCATGATGATGAGATTACACGTCGATGAAGCTGAATCTTGGTCACTGGAC 23458	
QY	974 CGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGGATCAATTTGAACGCTCTGCAGCTC 1033	
Db	23457 CAGAGAGAGCGCAGAGGCTTATGAGCTTTAGAGCTGAGACCCACGCTCAGTAGCTC 23398	
QY	1034 GAGCTAATCAAGAGAGCTGAATTCGCTAGAGAAAGAGGCGGATGTCGAGTAGATATG 10939	
Db	23397 GAGCGAATGAGAAGAGAGCTGAGATTGCTAGA-GAAAGAGAGCAATGACACGAGATATG 23339	
QY	1094 AGCTGATTGATCAGATATCAAACTGAGTATGAGCCAGAGCTCATGGCGAAGAGAGCA 1153	
Db	23338 AGTTGATCGACGAAGATATTGACGTCGAGTATGAGCCTGAGTCAATGGCAGAGAAACGA 23279	
QY	1154 AGCTACTGAACAAATCCGACGAGGTTACAGTGGAGGAGTATATCAGATTCTTTGAGATGA 1213	
Db	23278 ACCTGTTGAACAAGCT-----GATGA 23257	
QY	1214 ATGACTTCTGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCAGTTGGGTTACTGG 1273	

Thu Sep 9 11:17:24 2004

us-10-088-384a-27.rng

Db 34871 TCATCAAGTGTTCAGAGCCATCAGTTCTGACAGACAAGCTTAGCTGCTCTCTCTA 34812
Qy 2383 CTACAACTATTTCGCAATGACAGCTCCCTCAGGACATGCCCTTCAGAGGATATGAGC- 2440
Db 34811 CCACAGCTATTTCGCGAGGAGAGGCTCCTCAGGACATGCCCTTCAGAGGATATGAGC- 34752
Qy 2441 -----GCCGAGCTTAGAGACAGA 2460
Db 34751 CAGGCCAAGTCATCACAGGCTGAGCCAGTCACCAAGCTGAGCTAGTGACCGAG 34692
Qy 2461 AGATTCTGCTATGCTCCGCGAGGCTCATCTTCAGGCTCGTGAATCTAGGAGAAATA 2520
Db 34691 TAGTCCACCAAGTCCCTCAAGGCTCATCTTCAGGCTCGGAGCTCGGAGAAAGA 34632
Qy 2521 GGAGAACACACTACTCTGATCTAGCAGGAGCAGACGACTCTTCGAGTCTGCTAGT 2580
Db 34631 AGAAGGCTGCATCTGCTCGGTTCGCGAGGAGTACAGACTCTTCGAGTCCCTGAGT 34572
Qy 2581 TACGCGACCGGCTGCTCGGCGCAATAGAGAGAGAGTTCGAGTATCTTCAGAGCGGTG 2640
Db 34571 TACGCGACCGGCTGCTCGGCGCAATAGAGAGAGAGTTCGAGTATCTTCAGAGCGGTG 34512
Qy 2641 CTGCGCGCACAGAGCTGATGAGATCGATACCCACATGCTGAGTATGATCGGACATG 2700
Db 34511 CTGCGCGCACAGAGCTGATGAGTCCGAGTACCCCGAGGGAAGTTCGAGACACAGCAG 34452
Qy 2701 GCGGTGCTGATGCTGCGGAGCATCACAGGAGCCTTCACTACCAACTTCGTTCTAT 2760
Db 34451 GAGATTCTTCGATGCTCGGAGCAATCACAGGAGCTATTCGCGACCAACTTCGTTCT 34392
Qy 2761 TATTCAGTGTAGTAAAGCGCTTCACTTCACTATATATATATATATATATATATAT 2820
Db 34391 TCTTCACATGAGTATGAGCTTCACTTCACTATATATATATATATATATATATATAT 34340
Qy 2821 TCTTTATTTTGTTCAGTGAATGGAATTCCTGCTGAGTATCTCTTCCAAAGTTTATTCAC 2880
Db 34339 TTTTATTTTGTTCGATGTTTGTCTTAAGTATCTCTTCCAAAGTTTATTCAC 34280
Qy 2881 CAGTGTGCTGATTTAAGTTTGGGAGGCTCAGGAGTATG- -TTGCAATCTATA 2938
Db 34279 CAGTGTGCTGATTTAAGTTTGGGAGGCTCAGGAGTATG- -TTGCAATCTATA 34220
Qy 2939 TATTTTAAAGTCTGATTCATCTAAGGCTAGAGAA- -ACCAAAAAAATTTAAATTTTC 2997
Db 34219 TAACTTGTGCTGATTCATCTAAGGCTAGAGAA- -ACCAAAAAAATTTAAATTTTC 34160
Qy 2998 AGAAATGATTTCAAAAAAGAGTGTTCATGTAGTTCATTTAGGATCAAGTC 3057
Db 34159 AGAAATGATTTCAAAAAAGAGTGTTCATGTAGTTCATTTAGGATCAAGTC 34100
Qy 3058 TAGAGTGTTCATTTAGGATTTGTCATATGATAGGGAATATGATAGATAGCCTTGT 3117
Db 34099 TAGAGTGTTCATTTAGGATTTGTCATATGATAGGGAATATGATAGATAGCCTTGT 34040
Qy 3118 AAGCA 3122
Db 34039 AAGCA 34035

RESULT 12
AAF30450
ID AAF30450 standard; cDNA; 2512 BP.
XX AC AAF30450;
XX DT
XX 29-MAY-2001 (first entry)
XX Transcriptionally silent information clone TSI-A 5' extension.
XX TSI-A; transcriptionally silent information; gene silencing; ss.
XX Arabidopsis thaliana.
OS

XX FH Location/Qualifiers
FT CDS 437..2383
FT /*tag= a
FT /transl_except= (pos:1631..1633, Xaa)
FT /note= "Xaa = stop codon"
XX
XX WO200120010-A2.
XX
XX 22-MAR-2001.
XX
XX 14-SEP-2000; 2000WO-EP008994.
XX
XX 16-SEP-1999; 99GB-00021964.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Steimer A, Mittelsten Scheid O, Paszkowski J;
XX
XX WPI; 2001-244815/25.
XX P-PSDB; AAB20321.
XX
XX Novel transcriptionally silenced genes obtained from Arabidopsis useful
XX for selecting plants which compared to wild type plant are impaired in
XX transcriptional gene silencing.
XX
XX Claim 1; Page 21-22; 36pp; English.
XX
XX This present sequence is that of a 5' extension of an Arabidopsis
XX thaliana transcriptionally silent information (TSI) sequence, i.e. TSI-A.
XX The TSI-A clone was identified by differential mRNA screening. Comparison
XX of transcriptional gene expression between an Arabidopsis line carrying a
XX silent transgene present in multiple copies and its mutant derivative
XX mom1 impaired in silencing of the transgene revealed 2 cDNA clones which
XX were expressed in the mutant plants, but not in the parental or wild-type
XX plants. Both clones were derived from the same family of transcripts
XX referred to as TSI. Genomic templates encoding TSI are repetitive
XX elements with mainly pericentromeric location and conserved organisation
XX among various ecotypes. Transcriptional silencing of genomic TSI
XX templates is specifically released in the mutant. Transcription of TSI
XX can be used as a marker to identify a defective silencing pathway in a
XX plant. The present 5' extension of TSI-A was obtained by PCR. The
XX sequence includes a putative open reading frame for a 648 amino acid
XX protein (see Aab20321), but includes an in-frame stop codon. It shows 97%
XX identity to another isolated 5' extension sequence (see AAF30451). The
XX invention provides a nucleic acid encoding genetic information which is
XX silenced in wild-type plants but whose expression is turned on in plants
XX defective in transcriptional gene silencing, and has the formula: RA-RB-
XX RC, where RA, RB, RC constitute component nucleotide sequences selected
XX from the present sequence and those given in AAF30450-59. Also provided
XX are methods of selecting a plant that is impaired in transcriptional gene
XX silencing, and of producing DNA representing at least part of a gene
XX necessary to maintain silencing of another gene in a cell or plant
XX
XX Sequence 2512 BP; 640 A; 616 C; 634 G; 622 T; 0 U; 0 Other;

Query Match 39.4%; Score 1230.8; DB 4; Length 2512;
Best Local Similarity 74.9%; Pred. NO. 0;
Matches 1644; Conservative 0; Mismatches 467; Indels 83; Gaps 5;
Qy 788 ACTCAACCGTTTACTCGACCGGTACGCTGGTTAGATTGTTTGGTTTGGTTTCTATTA 847
Db 328 ACTCGACCGTTTCACTCGATCGCGTACTTGACT--GGTTAGTGTGTGTTTATTGAA 385
Qy 848 CTAACTATTTAACGTTTATCTTTGAGTTTC-GTCTGTTTGTAGTTTCATCATGTAAC 906
Db 386 CTAACATATTGATATTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTT 445
Qy 907 TAGAGTGAATAATCTCTATGACCGCTGATTATATGATGATGATGATGATGATGATGAT 966
Db 446 TACAGTGGCAATCTCTCATGATGCGGATTTACACGTCGATGATGATGATGATGATGAT 505
Qy 967 ACTAGACCGGAGTGAAGACCAATGTTTACGAGAGCTATAGGGAATGATTTGAAGCTCT 1026

QY 2262 AGCGAAGCTCACAGAACACAGCAAGCTGCAGAGTGGTGCAGAACACAGATAAGTT 2320
DB |||||||
25897 AGCGAAGCTCATAGAAAACACAGCAAGCTGCAGAGTGGTGCAGAACACATGATAAGTT 25839
RESULT 14
AAE22290/c
ID AAE22290 standard; DNA; 96988 BP.
XX AC . AAE22290;
XX XX
XX 20-MAR-2001 (first entry)
XX DE BAC containing repeats from centromeres 1-4 #13.
XX XX Centromere; michrosome; vector; ds.
XX XX Arabidopsis thaliana.
XX OS
XX PN WO200055325-A2.
XX XX
XX PD 21-SEP-2000.
XX XX
XX PF 17-MAR-2000; 2000WO-US007392.
XX XX
XX PR 18-MAR-1999; 99US-0125219P.
XX PR 01-APR-1999; 99US-0127409P.
XX PR 18-MAY-1999; 99US-0134770P.
XX PR 13-SEP-1999; 99US-0153584P.
XX PR 17-SEP-1999; 99US-0154603P.
XX PR 16-DEC-1999; 99US-0172493P.
XX XX
XX PA (UYCH-) UNIV CHICAGO.
XX XX
XX PI Preuss D, Copenhaver G, Keith K;
XX XX WPI; 2000-587529/55.
XX XX
XX DR
XX XX
XX PT Recombinant DNA construct comprising a plant centromere, useful for
XX PT producing stably inherited michrosomes which can serve as vectors for the
XX PT construction of transgenic plant and animal cells.
XX PS Claim 102; Page 550-572; 1449pp; English.
XX XX
XX CC The present invention relates to a recombinant DNA construct of a plant
XX CC (Arabidopsis thaliana) centromere. The constructs are useful for
XX CC producing stably inherited michrosomes which can serve as vectors for the
XX CC construction of transgenic plant and animal cells expressing selected
XX CC proteins such as hormones, enzymes, interleukins, clotting factors,
XX CC cytokines, antibodies, and growth factors
XX SQ Sequence 96988 BP; 29654 A; 18289 C; 18523 G; 30522 T; 0 U; 0 Other;
Query Match 37.6%; Score 1174.8; DB 3; Length 96988;
Best Local Similarity 70.8%; Pred. No. 6.3e-304;
Matches 1991; Conservative 0; Mismatches 237; Indels 584; Gaps 9;
QY 373 ACTTATCACTTCAAGCTCCTCATCTCTTCATCGTTTCCAAACCCGCTGCTCTCATCCCC 432
DB |||||||
68164 ACCTTACCACCTCTATCTCCAAAGATAACACTCGACCTCGCGCTTCTCACCGCTCTCC 58105
QY 433 CACGAAGCTTGTTCATCACTCTCATCTCATCCAGTTTCATCGATTCAGCAACCAACT 492
DB |||||||
68104 ATCGTCACCGCCTGCTCACTCTCTCCAAAGGAACAACCTCCAGCTCTCCATTTCACTCACT 68045
QY 493 CGACCTCGTCTCTTTGCCACTCATAGTCACTCGATCTCTCTCCACCATCTTCATCATCT 552
DB |||||||
68044 TGACCTCGTCTCTCTCGCTCTCACAATCACTCGATCTCTCTTCAACCATCTTCATCACT 67985
QY 553 CCCTTACTCGACACCGTGGTCTCGCTCCACCATGCCATTAAAGCTCACTCGATT- 611
DB |||||||
67984 ACCATCTCGACACCGTGGTCTCGCTCCACCATGCCATTAAATCTCACTCGATTG 67925

QY 612 ----GTCAAAGAGAGAGAGTGAAGCTCAACACCGCACTCGACCGGTTTCCCTCTA 667
DB |||||||
67924 CCGAGTCAAGAGAGAGAGTGAAGCTCCATCAACGCCACTCGACCTTGTTCCTCTA 67865
QY 668 CAGATTCAACACTCGACCAAGTCTACCATTCTCCACACCGCTCTTTTCCCATACAC 727
DB |||||||
67864 CACATTC AAC----- 67855
QY 728 TCGACCAACAACACTCTCAAGTAAAAAAGAAAAAAGTCAAAACCGACAGTTTC 787
DB |||||||
67854 ----- 67855
QY 788 ACTCAACCGGTTTACTCGACCGGTACGCTGGTTAGATTGTGTTTGGTTTGTCTATTA 847
DB |||||||
67854 -----ACTCGACCGGTACAGTGGTTTGGTGTGCTTTTGGTTTGTCTGTTA 67808
QY 848 CTAACATATTAAAGTTTATCTTTAGTTTCTGTTCTGTTTGTAGTTTTCATCATAGTAAC 907
DB |||||||
67807 CTAACGCAATTAAAGTTTATCTTTAGTTTCTGTTTTCAGGTTTTCATCATAGTAAC 67748
QY 908 ACAGTGGAAAAATCCTCTATGACCCCTGATTATATGTTGGATGAGCTAAAGTCTGTGTTCCA 967
DB |||||||
67747 ACAGTGGAGAAATCTTATGACCCCTGATTATATGTTGGATGAGCTGAGTCTGTGTTCCG 67688
QY 968 CTAGACCGGAGTGAGAGCAACATGTTTACGAGAGCTATAGGGATGAATTTGAACGCTCTG 1027
DB |||||||
67687 CTAGACCGAGG---GAGCAACATGTTTACGAGAGCTTTAGGGATGAATTTGAACGCTCCG 67631
QY 1028 CAGCTCGACGTAATCAAGAGAGCTGAATTCGCTGAGAGAAAGAGGGCGATGTCGAGTA 1087
DB |||||||
67630 CAGCTCGACGTAATCAAGAGAGCTGAATTCGCTGAGAGAAAGAGGGCGATGTCGAGTA 67571
QY 1088 GATATGAGCTGATTGATGAGGATATCAAACTCAGTATGAGCTGAGCTGAGTCAAGGCAAG 1147
DB |||||||
67570 GATATGAGCTGACTGATAGGATATCGAACTGAGTATGAGCTGAGCTGAGTCAAGGCAAG 67511
QY 1148 AGACGAAGCTTACTGAACAAATCCGACGAGTTACAGTGGAGGAGTATATCAGATTCTTTG 1207
DB |||||||
67510 AAACGAGGCTGTTGAACAAACCGATGAGTTACAGTGGAGGAGTATATCAGATTCTTTG 67451
QY 1208 AGATGAATGACTTCTGGGAAACGAGGTATCCCTGATATGAGACTTTAGCCAGTGGGGT 1267
DB |||||||
67450 AGATGAATGACTTCTAGGAAACGAGGTATCTTTATTATGAGACTTTAGCCAGCTCGGT 67391
QY 1268 TACTGAGACGTCGACGATCTGTTTCGAGAGTGTCTATCTG----- 1308
DB |||||||
67390 TACTGAGACGTCGACGATCTGTTTCGAGAGTGTCTATCTGAGACACTCATGTCTTACC 67331
QY 1309 -----ATAAGGAGGAGACAATCGAGTTTCTTTCCACACTGCAAGTGGAAATGT 1356
DB |||||||
67330 CTTACGCTCTCTATTAAGGAGGAGACAATCGAGTTTCTTTCCAGCTGCAAGTGGAGATGT 67271
QY 1357 ATGAGGAGTCAAGACTTTGAGCTGGATACCATGGGGTTAGGCTTTTGAAGTCTTTAG 1416
DB |||||||
67270 ATGAGGAGTCAAGACTTTGAGCTGGATACCATGGGGTTAGGCTTTTGGGGTTCGCCAG 67211
QY 1417 TCGATGAACAGCGGTACCAAGATTTAGATCAAGAAATTCGAAGACTGTTTGGTTTCCCTA 1476
DB |||||||
67210 TCGATGAACAGCGGTATCAGCTTTGATCAAGAGTTGGAAGAACTGTTTGGTTTCCCTA 67151
QY 1477 GTGGAAGAGAAACCAACCCAGGTTTGACAGGGAAGAGCTTAAGGATTGTTGGGCTACTA 1536
DB |||||||
67150 GTGGAAGAGAAACCAACCCAGATTGATAGAGA----- 67117
QY 1537 TTGGGAACAATCTACCGCTAAACTCGACCGGTCCAAGAGCAACCAAAATCCGAGTCTG 1596
DB |||||||
67116 ----- 67117
QY 1597 TGATTCGCTACTTTTCAGCGCTCGGTGCAATGTTTTTACTCCAGGGAGTCTACAGGA 1656
DB |||||||
67116 ----- 67117
QY 1657 CGGTGCTTAACACAGACATGAAGATGATAGATTTCAGCGCTTATAGGGATTCTCCGCTTA 1716

Db	67116	-----	67117	
Qy	1717	CAAAAGGAAAGAAATGTCCTGAGAGGAGATCTTAACGACTCACCACAGTAATGCCCTCTGT	1776	
Db	67116	-----	67117	
Qy	1777	TGATCCATCTGTGTGGGTACAATGAATGGCGGCTGACAAACGGCAAGAAAGGTAAAGAG	1836	
Db	67116	-----	67117	
Qy	1837	GAGCAGTATGCGTGGTGGCGTTGTGACGCCAATTTCTGAAAGTTTGTGGAGTTCCCGCTCA	1896	
Db	67116	-----	67117	
Qy	1897	AGGAAGTAGGTTAGCACCAGAAATGATGGAATTTGGATCACTTGGCCCGGATGTGAGTTCT	1956	
Db	67103	AGGAACAGGGTTAGCACCAGGAGATGATGAGCTTTGGATCACTTGGCCCGGATGTGAATCT	67044	
Qy	1957	CTGAGTTTGACATGTTGGCGACATTTACCGCTACAGGTTTCGAGCTTCATTCAGATAGAA	2016	
Db	67043	TGGAGTTTGACAT-----GGTTTCGAGCATTCATCGTTTAAAG	67007	
Qy	2017	TCGCCAATCTTTTCCCTGCTATTTACGCTACTAGGATTTCTCGAGGCGAGGAACATTTG	2076	
Db	67006	TCGCCAATCTTTTCCCTGCTATTTACGCTACTAGGATTTCTCGAGGCGAGGAACATTTG	65947	
Qy	2077	ACTTCAAGCCTGCGCTTGAAGATCTTTATTTTCGAGGCGAGTCCGCCCACTGAGGAGATTA	2136	
Db	65946	ACTTCAAGCCTGCGCTTGAAGATCTTTATTTTCGAGGCGAGTCCGCCCACTGAGGAGATTA	66887	
Qy	2137	GTCAACCGAGGAGCTCAATAGAGATGTTGATGACACATATGATATGATGATGAGCGG	2196	
Db	66886	GTCAACCGAGGAGCTCAATAGAGATGTTGATGACACATATGATATGATGATGAGCGG	66827	
Qy	2197	AGTTTGACACGAGCATGATCATTTTCAGTGAGCATATACCTCCAGCGAGGAAAGCAAGA	2256	
Db	66826	AGTTTGACACGAGCATGATCATTTTCAGTGAGCATATACCTCCAGCGAGGAAAGCAAGA	66767	
Qy	2257	GTTTGAGCGAAGCTCACAGGAACAACAGCAAGCTGCAAGTGGTGCAAGAAAAGAGTATG	2316	
Db	66766	GTTTGAGCGAAGCTCACAGGAACAACAGCAAGCTGCAAGTGGTGCAAGAAAAGAGTATG	66707	
Qy	2317	AGTTACTGCGAAGTGCCTTCAGGCTATCAAGTTTCTGAAGCAAGATCAGCTGCTCCT	2376	
Db	66706	AGTTACTGCGAAGTGCCTTCAGGCTATCAAGTTTCTGAAGCAAGATCAGCTGCTCCT	66647	
Qy	2377	CTTCCACTACAACTATTCGCAATGACAGCTCCCTCAGGACATGCTTTCGAGGAGATATG	2436	
Db	66646	CTTCCACTACAACTATTCGCAATGACAGCTCCCTCAGGACATGCTTTCGAGGAGATATG	66587	
Qy	2437	ACGCGCCCGAGCC-----TAGAGAGCAAGAGATTCGCTATGCC	2475	
Db	66586	ACGCGCCCGAGCCAGTTCGCGCGAGGCTTAAAGCTTAGAGAGCAGAGATTCGCTATGCC	66527	
Qy	2476	CTGCGAGGATTCATCATTTTCGAGGCTTCGTAATCTAGGAAGAAATAGGAGAAACACATCA	2535	
Db	66526	CTGCGAGGATTCATCATTTTCGAGGCTTCGTAATCTAGGAAGAAATAGGAGAAACACATCA	66467	
Qy	2536	CTCGATCTAGCAGCAGGAGCAGACGACTTCTGAGTCTGATGTTTACCGGACCCGGTG	2595	
Db	66466	CTCGATCTAGCAGCAGGAGCAGACGACTTCTGAGTCTGATGTTTACCGGACCCGGTG	66407	
Qy	2596	CTGCGCCCAATAGAAAGAGGCTCGATATCTTCAGAGCGGTGCTGCCGCCACACAGAG	2655	
Db	66406	CTGCGCCCAATAGAAAGAGGCTCGATATCTTCAGAGCGGTGCTGCCGCCACACAGAG	66347	
Qy	2656	CTGATGAGATCGAGTACCCACATGCTGGAGCTGATACGGAACATGCGGTTCTGCTATGG	2715	
Db	66346	CTGATGAGATCGAGTACCCACATGCTGGAGCTGATACGGAACATGCGGTTCTGCTATGG	66287	
Qy	2716	CTTGGGAGCAATCACAGGAGCCATGACTACCAATCTGTTCAATTTTCGACTGAGGTA	2775	

Qy	2828	TTTCTTTCAGTGGATTGTTGCTCTGAGTACTCTCTTCCAAAGTTTATTTCACACAGTGGG	2887
Db	1918	TTTGTGTTTGTGATGTTGTTTGTCTCTGAGTACTCTCTTCCAAATTTGTTCTCACACAGTGGG	1977
Qy	2888	CTGTGTGATTTAAGTTTGGG	2907
Db	1978	CTGTGTGATTTAAGTTTGGG	1997

Search completed: September 9, 2004, 04:01:20
Job time : 1136 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2004, 03:36:11 ; Search time 239 Seconds
(without alignments)
7249.191 Million cell updates/sec

Title: US-10-088-384A-27
Perfect score: 3122
Sequence: 1 ttcataattcgacctcttc.....atgagatagcctgtgaagca 3122

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751.6	24.1	1857	US-09-322-478-24	Sequence 24, Appl
2	90	2.9	7218	US-08-232-463-14	Sequence 14, Appl
3	63.8	2.0	1166	US-09-072-596-323	Sequence 323, Appl
4	63.8	2.0	1166	US-09-072-596-328	Sequence 328, Appl
5	63.8	2.0	7218	US-08-232-463-14	Sequence 14, Appl
6	55	1.8	53526	US-08-658-136-2	Sequence 2, Appl
7	55	1.8	53577	US-08-658-136-1	Sequence 1, Appl
8	45.4	1.5	6583	US-10-204-708-25	Sequence 25, Appl
9	42.6	1.4	318	US-09-165-264-12	Sequence 12, Appl
10	41.6	1.3	32207	US-08-770-379-20	Sequence 20, Appl
11	41.6	1.3	32207	US-08-757-669A-20	Sequence 20, Appl
12	41.6	1.3	32207	US-09-230-371A-20	Sequence 20, Appl
13	40.6	1.3	832	US-09-621-976-2813	Sequence 2813, Ap
14	40.4	1.3	487	US-09-257-179-22	Sequence 22, Appl
15	40.2	1.3	320	US-09-165-264-7	Sequence 7, Appl
16	40	1.3	320	US-09-165-264-11	Sequence 11, Appl
17	39.8	1.3	289	US-09-007-005-17	Sequence 17, Appl
18	39.8	1.3	289	US-09-244-796-17	Sequence 17, Appl
19	39.6	1.3	320	US-09-165-264-14	Sequence 14, Appl
20	39.6	1.3	12001	US-08-458-568A-11	Sequence 11, Appl
21	39.4	1.3	5852	US-07-867-106-2	Sequence 2, Appl
22	39	1.2	867	US-09-482-273-50	Sequence 50, Appl
23	39	1.2	3182	US-08-971-395-1	Sequence 1, Appl
24	39	1.2	3182	US-08-413-135-1	Sequence 1, Appl
25	38.8	1.2	832	US-09-621-976-2813	Sequence 2813, Ap
26	38.6	1.2	470	US-09-020-956-102	Sequence 102, App
27	38.6	1.2	470	US-09-030-607-102	Sequence 102, App

Sequence 102, App
Sequence 102, App
Sequence 102, App
Sequence 102, App
Sequence 102, App
Sequence 102, App
Sequence 102, App
Sequence 102, App
Sequence 80, Appl
Sequence 23, Appl
Sequence 195, App
Sequence 81, Appl
Sequence 8, Appl
Sequence 13, Appl
Sequence 1, Appl
Sequence 11, Appl
Sequence 6, Appl
Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-322-478-24
; Sequence 24, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-322-478-24

Query Match 24.1%; Score 751.6; DB 4; Length 1857;
Best Local Similarity 65.3%; Pred. No. 1.9e-210;
Matches 1250; Conservative 0; Mismatches 564; Indels 100; Gaps 6;
QY 898 ATGAGTAACACAGTGGAAATCCTTATGACCCCTGATTATAATGTGATGAAGTAAG 957
DB 1 ATGAGCAATTACAGTGGCAGTTCTTCTGTGATCCTGACTACACATGGATGAGACAGAA 60
QY 958 TCCTGGTCACTAGACCGGAGTGAGAGCAACATGTTTACGAGGCTATAGGATGAATTT 1017
DB 61 TCGTCATCTTCAAGGCCAGAGAGACAGAGAGAAATACGAAAGTTTCAG----- 110
QY 1018 GAACGCTGCGAGCTCGACGTAATCAAGAGAGCTGAATCGTAGAGGAAGAGGGCG 1077
DB 111 -----AAGGAAAGCTGAGATAGCCCGAGAAAGAGAGCG 144
QY 1078 ATGTCGCTAGATATGAGCTGATTGATGAGGATATCAAAACCTGAGTATGAGCCAGATCA 1137
DB 145 ATGACGAGAGGTATGAGCTTATAGACGAGATCTGGAGGAGGATATGCTCTGAAACAG 204
QY 1138 TGGCCCAAGGAGACGAACCTACTGAAACAAATCCGACGAGGTTTACAGTGGAGGATATATC 1197
DB 205 ACTGCGAGAGCTACCAAACTTTCGCAACGCGGACATTTGCGCTGCTGAGGAATATGTT 264
QY 1198 AGATCTTTGAGTGAATGACTCTCTGGGGAACGAGGATATCCTGTATATGAGACTTTAGCC 1257
DB 265 AGGCTTTTCAAGCTGAATGAGTTCTGTAGCAGAGGTATCTCTGCTCGACCTCACTTGA 324
QY 1258 CAGTTGGGGTTTACTGGAGGACGTCGAGCATCTGTTTCGAGAAAGTGTGTCATCTG----- 1308

Db 325 CAACCTGGATTGTTGGAGATGTTTCAGACCTGTACCAAAAGTTGTCATCTGGACACTTTG 384
Qy 1309 -----ATAAGGAGGAGCAATCGAGTTCTTTCCACACTGCAA 1346
Db 385 ATGGCTTTATCGGTATGTAGCATATGAAAGATGAGACAAATACCAATTCCTCTCCACACTCAA 444
Qy 1347 GTGGAATATAGAGGACTCACAGACTTTGAGCTGGATACCAATGGGGTTAGGCTTTCTTG 1406
Db 445 GTAGAGCTCTACCAAGGTATGACCTCTGATGAGTTGGAATGTGAAGNATTGGGATTTCTG 504
Qy 1407 ACGTTCCTTAGTGATGAACAGCGGTACAGATTTTAGATCAAGAATTTGGAAGAACTGTTT 1466
Db 505 CGATTTCTGTGTATGTTGTCATGATGATGAGTTGGAATTTGGAAGGATTTGTTT 564
Qy 1467 GGTTCCTAGTGAAGAGGAAACCAACCCAGGTTTGACAGGGAGAGCTTAAGGATTTG 1526
Db 565 GAITTTCCAGTGAACGGGATCTAAGCCAAAGTATGAAGAGAGAGTTGAAAGACTTG 624
Qy 1527 TGGGCTACTATTGGGAACAATCTACCGCTAAACTCGACGGCTCAAGAGCAACCAATC 1586
Db 625 TGGATCACATCGGCAGCTCTGTACCGTTGATGCTTCCAGGTCAAAGAGCAATCAGATA 684
Qy 1587 CGGAGTCTGTGATTCGCTACTTTTCAGCTCGGTGCGAATGTTTTTACTCCAGGAG 1646
Db 685 CGCAGCCCTGTCTCATCAGGTACTTCCAGCGTTCTGTAGCAAGTACTCTACTCCCGAGAG 744
Qy 1647 TCTACAGCACCCTGTCTAAACACAGACATGAGATGATGATTCAGCGCTTATAGGAT 1706
Db 745 ATTACAGGACTGTCTCACTAACTCTGATATGAGATGATCGCAATGGCCCTCAAGGAAC 804
Qy 1707 CTCGCCCTTACAAAAGGAAAGAAATCTCCTGAGGAGAGATCTTAACGACTCACCAACAGTA 1766
Db 805 CTCGCCCAACTTAAATGCGATGTCCTCCAGGGTGAATCAATGACACACTCTCTCT 864
Qy 1767 ATGCTCTGTGATCCATCTGTGTGGGTACATGAAGTGGCGCTGACAAAACGGGAAGAG 1826
Db 865 ATACTTCTTCTGATCCATCTGTGTGATACAAAAACTGGCGGTGAGCAATTAACCGCAAG 924
Qy 1827 AAGTAAAGGAGGAGCACTAGCTGCTGGGTGGCTGTGAGCCCAATTTCTGAAGTTTGTGA 1886
Db 925 AGAGCAGCAGGCGCTCTGTGCAATGAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 984
Qy 1887 GTTCGCTCAAGGAAGTAGGGTTAGCACCGAGAATGATGACTTGGATCACTTGGCCGCA 1946
Db 985 GTCCCACTCATTTCTGCTGACTCGAGCCACGAGCAATGATATCGACACTACGTCTAC 1044
Qy 1947 TGTGATCTCTGAGTTTGACATGTTGGCGATTTCCCGCTACCGCTACAGGTTCCAGCATTTCA 2006
Db 1045 TGCCAAATCTCGAGTTTCAATGGTTGACGATTTCCACAGGTTTCAAGTTTGGCACTCT 1104
Qy 2007 TGATTAGAATGCCAACAATCTTTTCCCTGCATTTAGCTACTAGGATTTCTCGAGGCG 2066
Db 1105 ACAGCAGGAGAGCTAACTCTCTCCCTAGCCCTGAGTCAACGATATCGAGGA 1164
Qy 2067 AGGAACATTTGACTTCAAGCTCGCTTGAAGATCTTTATTTGAGGGCAGTCCGCAACT 2126
Db 1165 GATAACATTTAGTCTGAGTTGAGGCTCTACTATGAGAACGCTCCACCATTA 1224
Qy 2127 GAGGAGATTAGTCAACCGAGAGCTCAATAGAAGATTTGATGAGACATATGATATA 2186
Db 1225 GATGAGGAGATCTTCTTGAAGAAGCT-----GCTTCGATGGGATGATGAAGAT 1275
Qy 2187 GATGAGGAGGATTTCACAGCAGCATGATCATTTTCAGTGAGCATATCTCCAGCGAGG 2246
Db 1276 GGAGCAGTAAGTTTCGACACTACATGATATCATTTTGTGTAACATGTACTCCAGCGAGG 1335
Qy 2247 AAAAGCAAGATTTGAGCGAGCTCAAGGAAACACAGCAGCTCGAAGTGGTCAAG 2306
Db 1336 CAGAGCAAGAGCTTGAAGCTCATAGAAATTAAGTAATTAAGTAATTAAGTAATTAAGTA 1395
Qy 2307 AAACAGATAGTTACTCGCCCAAGTGCCTCAGGCTATCAAGTTTCTGAGGACAGATC 2366

Db 1396 AGCAGAGCAGGCTGATCGCCAAGTG-----TTTCAAGCTTCTGACAGACAAGCTG 1446
Qy 2367 AGCTGCTCTCTTCCACTACAAATTTCCGCAATGACAGCTCCCTCAGGACATCCCTTCG 2426
Db 1447 AGTTGCTCTTCTCCCACTGCTATTTCCACAGGTACAACCTCTCTATGGAATGCCATCG 1506
Qy 2427 AGGAGATATGAGC-----GCCGAGCCTAGAGAGAGAAATTTCTGCATGTC 2474
Db 1507 AGGAGAATTAATGCACTTGGCAGCAGGCTGAGCTTAGCGAGCAGAGAGTCCCAATGTC 1566
Qy 2475 CTTGCGAGGATTCATCATTTGAGCCTCGTGAATCTAGGAAGAATAGGAGAACGACATC 2534
Db 1567 CAGCTAGGCAATTCCTGATTCGATCCGCGGAACACAAGAGAAAGAGGCTACACTC 1626
Qy 2535 ACTGATCTAGCAGCAGGAGCAGACTTCTGAGTCTCTGATTTACGCGACCGCGT 2594
Db 1627 ACTGATCTAGCAGC-----AGATCAGGCTCATTTCACTCGAGGAGATCACTCGACCGTGGT 1683
Qy 2595 GCTGGCGCAATAGAGAGAGGTCGAGTATCTCAGAGCGGTGCTGGCGCCACACAGA 2654
Db 1684 GCTGGCGCAGCAGAGGAGAGATGTCGAGTTCTCAGAGCGGTGCTGGCGCCACACAGA 1743
Qy 2655 GCTGATGAGATCGAGTACCCACATGCTGAGCTGATACGGAACATGGCGGTTGCTGCTATG 2714
Db 1744 GCTGATGAGTCTGAGTACCCATCTGCTGAGCTGATACAGAAACAGGAGGTTGCTGCTATG 1803
Qy 2715 GCTTGGGAGCAATCAGAGCAGGAGCAGTACCTACCACTTCTGCTCATTTTCGAC 2768
Db 1804 GCCTGGGAGCAATCGAGCAGCAATTGACGAGCAACTACGTTCTTCTTCTGAC 1857

RESULT 2

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEFAX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs

STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 53577 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

```

1  ; GENERAL INFORMATION:
2  ;
3  ; APPLICANT: OLEK, Alexander
4  ; APPLICANT: PIENEBROCK, Christian
5  ; APPLICANT: BERLIN, Kurt
6  ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
7  ; TITLE OF INVENTION: by Assessing DNA Methylation
8  ;
9  ; FILE REFERENCE: 5013.1012
10 ;
11 ; CURRENT APPLICATION NUMBER: US/10/204,708
12 ;
13 ; CURRENT FILING DATE: 2003-05-06
14 ;
15 ; PRIOR APPLICATION NUMBER: PCT/EP01/033971
16 ;
17 ; PRIOR FILING DATE: 2001-04-06
18 ;
19 ; PRIOR APPLICATION NUMBER: DE 10019058.8
20 ;
21 ; PRIOR FILING DATE: 2000-04-06
22 ;
23 ; PRIOR APPLICATION NUMBER: DE 10019173.8
24 ;
25 ; PRIOR FILING DATE: 2000-04-07
26 ;
27 ; PRIOR APPLICATION NUMBER: DE 10032529.7
28 ;
29 ; PRIOR FILING DATE: 2000-06-30
30 ;
31 ; PRIOR APPLICATION NUMBER: DE 10043826.1
32 ;
33 ; PRIOR FILING DATE: 2000-09-01
34 ;
35 ; NUMBER OF SEQ ID NOS: 98
36 ;
37 ; SEQ ID NO 25
38 ;
39 ; LENGTH: 6583
40 ;
41 ; TYPE: DNA
42 ;
43 ; ORGANISM: Artificial Sequence
44 ;
45 ; FEATURE:
46 ;
47 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
48 ;
49 ; US-10-204-708-25

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2004, 03:26:27 ; Search time 7044 Seconds
(without alignments)
13235.339 Million cell updates/sec

Title: US-10-088-384A-27
Perfect score: 3122
Sequence: 1 ttcatatattcgacctcttc.....atgagatagccttgtaagca 3122

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	725.6	23.2	855	28	BH237235
C 2	694.4	21.9	809	28	BH237437
C 3	682.2	21.9	801	28	BH237501
C 4	682.2	21.9	817	28	AQ958311

5	663.4	21.2	781	28	BH244328
6	647.2	20.7	848	28	BH237552
7	622.6	19.9	813	28	BH244305
C 8	611.6	19.6	806	28	BH235546
C 9	604.6	19.4	681	28	BH244756
10	601	19.3	691	28	BH244796
11	598.6	19.2	1101	28	B08362
C 12	587.4	18.8	704	28	BH237544
C 13	585.6	18.8	792	28	BH237151
14	577.2	18.5	670	28	BH237213
15	576.2	18.5	756	28	BH244293
C 16	572.4	18.3	725	28	BH244787
C 17	568.2	18.2	713	28	BH237198
18	553.6	17.7	684	28	BH244310
19	553.4	17.7	663	28	BH235487
C 20	551.4	17.7	662	28	BH235467
C 21	550.6	17.6	637	28	BH235552
C 22	543.4	17.4	685	28	BH235356
C 23	541.2	17.3	645	28	BH235469
24	529	16.9	617	28	AQ958677
25	528	16.9	740	28	BH235520
26	526.6	16.9	682	28	BH244766
27	521.2	16.7	582	28	B77880
28	519.6	16.6	771	28	BH237289
29	509.8	16.3	630	28	BH235434
C 30	504.2	16.1	558	28	BH244805
C 31	499	16.0	571	28	BH235416
C 32	499	16.0	623	28	BH235383
33	498.2	16.0	621	28	BH244296
C 34	498	16.0	554	28	BH244834
35	497.2	15.9	638	28	BH244311
36	487.4	15.6	545	28	B67818
37	486.6	15.6	627	28	BH244334
38	486	15.6	650	28	BH235456
39	485.2	15.5	765	28	BH244841
40	484.4	15.5	638	28	BH235402
41	481	15.4	645	28	BH235572
C 42	477.4	15.3	723	28	BH243461
C 43	476.6	15.3	638	28	BH235471
C 44	476.2	15.3	583	28	BH235549
45	475	15.2	654	28	BH235355

ALIGNMENTS

RESULT 1
BH237235/c
LOCUS
DEFINITION BH237235 855 bp DNA linear GSS 13-NOV-2001
AUIB10TR AUI Arabidopsis thaliana genomic clone AUIB10, genomic survey sequence.
ACCESSION BH237235
VERSION BH237235.1 GI:16907606
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 855)
AUTHORS Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Utterback, T.V., Feldblyum, T.V. and Fraser, C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC T2P24
JOURNAL Unpublished (2001)
COMMENT Other GSSs: AUIB10TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 849.
Seq primer: TR

QY	Db	2334	CTCAGGCTATCAAGT	2349
QY	Db	16	CTCAGGCTATCAAGT	1
<p>Class: sheared ends.</p> <p>Location/Qualifiers</p> <p>1. .855</p> <p>/organism="Arabidopsis thaliana"</p> <p>/mol_type="genomic DNA"</p> <p>/strain="Columbia"</p> <p>/db_xref="taxon:3702"</p> <p>/clone="AUII10"</p> <p>/clone_lib="AUII"</p> <p>/note="Vector: PHOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into PHOS2 using BstXI linkers"</p>				
<p>ORIGIN</p> <p>Query Match 23.2%; Score 725.6; DB 28; Length 855;</p> <p>Best Local Similarity 91.2%; Pred. No. 5e-122;</p> <p>Matches 781; Conservative 0; Mismatches 74; Indels 1; Gaps 1;</p>				
QY	Db	1494	CCGAGTTTGACAGGAAGAGCTTAAGGATTTGTGGCTACTATTGGGAACAATCTACCG	1553
Db	855	CCGAGATTTACAGAGAAGAGCTTAAGGATTTGTGGCACTATATAGGAACAATCTACCG	796	
QY	1554	CTAAATCTCGACGGCTCCAGAGCAACCAATTCGGAGTCCCTGATTCGCTACTTTTCAG	1613	
Db	795	CTAAATCTCGCGCGCTTTAAGGCAACCAATTCGGAGTCCCTGATCTGCTACTTTTCAG	736	
QY	1614	CGCTCGTTTGCCAAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTAACACAGAC	1673	
Db	735	CGATCGTGCGCAACGCTTTTACTCCAGGAGTCTACAGACCGTGTCTAACACGGAC	676	
QY	1674	ATGAAGATGATAGATTCAGCGCTTATAGGATTTCCCGCTTACAAAGGAAAGATGTC	1733	
Db	675	ATGAGATGATAGATTTATGCGCTTTACAGGATTTCTCCGGGACGAGGAGGAGATGTC	616	
QY	1734	CTGAGAGAGATCTTTAACGACTCACCAACAGTAAATGCTCTGTGTATCCATCTGTGTGG	1793	
Db	615	TTGAGAGCGATCTCAACACGACCAACAGTAAATGCTCTGTGTATCCATCTGTGTGA	556	
QY	1794	TACATGAATGGGGCTGCAAAACGCAAGAAAGTAAAGGACATATGCGTGGT	1853	
Db	555	TACAGGAATGGGGCTTGACAAACGGAAGAAAGTGAAGAGACATATGCGTGGT	496	
QY	1854	GGCTGTGAGCCAAATCTGAAAGTTGTGGAGTTCCGCTCAAGGAAGTAGGTTAGCA	1913	
Db	495	GGCGTTATGAGCCAAATCTGGAAGCTTGTGAAGTACCGCTCAAGGAAGTAGGTTAGCA	436	
QY	1914	CCGAGAATGATGGAATCTGATTCCTGCGCGGATGAGTCTCTGAGTTGACATGTT	1973	
Db	435	CCGAGGATGATGGAATCTGATTCATTCGCGAGATGAGTCTCTGAGTTGACATGTT	376	
QY	1974	GGGACATTTACCGCTACAGGTTCCGACATTCATCGATTAGATTCGCCAACATCTTTTC	2033	
Db	375	GGTGATTTTACCGCTACAGGTTTCGAGATTCATCGATTAAATTCGCCAACATCTTTTC	316	
QY	2034	CCCTGCAATTTACGTTACTAGGATTTCTCGAGGCAAGGAACATGACTTCAAGCTCGGCTT	2093	
Db	315	CCCTGCAATTTACGTTACTAGGATTTCTCGAGGCAAGGAACATGACTTCAAGCTCGGCTT	256	
QY	2094	GAAATCTTTATTTTCGAGGCGAGTCCGCCACTGAGGAGATTAGTCAACCGAAGAGCT	2153	
Db	255	GAAAGAACATTTATTTTCGAGGCGAGTCCGCCAACTGAGGAGATTAGTCAACCGAAGCT	196	
QY	2154	ACAATPAGAAGATGTTGTAGACATATGATATAGATGAGCGGAGTTTGAACAGAGATG	2213	
Db	195	ACAACAGAAGATATTTGATGAGCGATGATATAGATGAGCGGAGTTTGAACAGAGATG	136	
QY	2214	TATCATTTTCACTGAGCATATCTCCAGGAGGAAAGCAAGATTTTGAAGAGCTCAC	2273	
Db	135	TATCATTTTCACTGAGCATATCTCCGAGGAGGAAAGCAAGATTTTGAAGAGCTCAC	76	
QY	2274	AGGAACAACAGCAAGCTGCAGAAGTGGTGCAGAAACAGGATAAGTTACTCGCCAAAGTC	2333	
Db	75	AGGAACAACCAAGCTGCAGAAGTGGTGCATGAAGCAAGGATAAGTTACTCGCCAAAGTC	17	

QY 1963 TTGACATGTTGGCGACTTTCACCGCTACAGGTTGAGCATTCATCGATTAGATCGCA 2022
Db 450 TTGACATGTTGGCGACTTTCACCGCTACAGGTTGAGCATTCATCGATTAGATCGCA 391
QY 2023 ACATTCCTTTCCCTCGCATTTACGCTACTAGGATTCGAGGCGAGGACATTTGACTTCA 2082
Db 390 ACATTCCTTTCCCTCGCATTTACGCTACTAGGATTCGAGGCGAGGACATTTGACTTCA 331
QY 2083 AGCCTGCGCTTGAAGATCTTTATTTTCGAGGCGAGTCCGCAACTGAGGAGATTAGTTCACA 2142
Db 330 AGCCTGCGCTTGAAGACTTTTATTTTCGAGGCTAGTCCGCAACTGAGGAGATCATGTCATA 271
QY 2143 CCGAAGGAGCTACATAGAGAATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTG 2202
Db 270 CCGAAGGAGCTACATAGAGAATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTG 211
QY 2203 ACAGGAGATGATCATTTTCAGTGAGCATATACCTCCAGCGAGGAAGCAAGATTGGA 2262
Db 210 ACAGGAGATGATCATTTTCAGTGAGCATATACCTCCAGCGAGGAAGCAAGATTGGA 151
QY 2263 GCGAAGCTCACAGGAACAACAGCAAGCTGCAAGAGTGGTGCAGGAACAAGGATAAGTTTAC 2322
Db 150 GCGAAGCTCACAGGAACAACAGCAAGCTGCAAGAGTGGTGCAGGAACAAGGATAAGTTTAC 91
QY 2323 TCGCCAGTGCCTCAGGCTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCA 2382
Db 90 TCGCCAGTGCCTCAGGCTATCAAGTTTCTGAAGGACAAGATCAGCTGCTCCTCTTCCA 31
QY 2383 CTACAACATTTCCGCAATGACAGCTCCCTC 2412
Db 30 CTACAACATTTCCGCAATGACAGCTCCCTC 1

RESULT 3
BH237501
LOCUS
DEFINITION AU11805TR AU11 Arabidopsis thaliana genomic clone AU11805, genomic survey sequence.
ACCESSION BH237501
VERSION BH237501.1 GI:16908045
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 801)
AUTHORS Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Utterback, T.V., Feldblyum, T.V. and Fraser, C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC T2P24
JOURNAL Unpublished (2001)
COMMENT Other GSSs: AU11805TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 849.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. .801
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AU11805"
/note="Vector: pPOS2; site 1: BstXI; 2-3 kb sheared BAC
DNA inserted into pPOS2 using BstXI linkers"

FEATURES
source
1. .801
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AU11805"
/note="Vector: pPOS2; site 1: BstXI; 2-3 kb sheared BAC
DNA inserted into pPOS2 using BstXI linkers"

RESULT 4
AQ958311
LOCUS
DEFINITION LERAW77TR LERA Arabidopsis thaliana genomic clone LERAW77, genomic survey sequence.
ACCESSION AQ958311
VERSION AQ958311.1 GI:6786012
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Query Match 21.9%; Score 682.2; DB 28; Length 801;
Best Local Similarity 90.9%; Pred. No. 46-114;
Matches 726; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1623 GCCAATGTTTTTACTCCAGGGAGTCTACAGGACCGTGTCTTAAACACAGACATGAAGATG 1682
Db 2 GCCAACGCTTTTACTCCAGGGAGTCTACAGACACCGTGTCTTAAACACAGACATGGAGATG 61
QY 1683 ATAGATTCAGCGCTTATAGGGATTTCTCGCCCTTACAAAGAGAAAGATGTCTTGAGAGGA 1742
Db 62 ATAGATTCAGCGCTTATAGGGATTTCTCGCCCTTACAAAGAGAAAGATGTCTTGAGAGGA 121
QY 1743 GATCTTAAAGCACTACCAACAGTATGCTCTGTGATCCATCTCTGTGGGTACATGAG 1802
Db 122 GATCTTAAAGCACTACCAACAGTATGCTCTGTGATCCATCTCTGTGGGTACATGAG 181
QY 1803 TGGGCGCTTGACAAACGGCAAGAGGTAAGAGGAGCACTATGCTGGGTGGCGCTTGTG 1862
Db 182 TGGGCGCTTGACAAACGGCAAGAGGTAAGAGGAGCACTATGCTGGGTGGCGCTTGTG 241
QY 1863 ACGCCAAATCTGAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGTTAGCCAGGAATG 1922
Db 242 ACGCCAAATCTGAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGTTAGCCAGGAATG 301
QY 1923 ATGACTTTGGATCACTTTCGCGCGATGTGAGTTCTCTGAGTTTGTGATGGTGGTGTG 1982
Db 302 ATGACTTTGGATCACTTTCGCGCGATGTGAGTTCTCTGAGTTTGTGATGGTGGTGTG 361
QY 1983 CACCGCTACAGGTTTCGAGCAATTCATCGATTAGAATCGCAACATTTCTTCCCTGCAAT 2042
Db 362 CACCGCTACAGGTTTCGAGCAATTCATCGATTAGAATCGCAACATTTCTTCCCTGCAAT 421
QY 2043 TACGCTACTAGGATTCGAGGCGAGGAACAATTGACTTCAAGCCTGCGCTTGAAGATCTT 2102
Db 422 AACGCTTACTAGGATTCGAGGCGAGGAACAATTGACTTCAAGCCTGCGCTTGAAGATCTT 481
QY 2103 TATTTCCAGGCGCTCCGCCAATCGAGAGATTAGTACACCGGAGGAGCTACATAGAA 2162
Db 482 TATTTCCAGGCGCTCCGCCAATCGAGAGATTAGTACACCGGAGGAGCTACATAGAA 541
QY 2163 GATGTTGATGACATATGATATAGATGAGGCGGAGTTTGAACGAGCATGTATCATTTTC 2222
Db 542 GATGTTGATGACATATGATATAGATGAGGCGGAGTTTGAACGAGCATGTATCATTTTC 601
QY 2223 AGTCAGCATATACCTCCAGCGAGGAAGCAAGAGTTTGAAGCGAGTTCACAGAAACAAC 2282
Db 602 AGTCAGCATATACCTCCAGCGAGGAAGCAAGAGTTTGAAGCGAGTTCACAGAAACAAC 661
QY 2283 AGCAAGCTGCAGAGTGTGCAAGAAACAGGATAAGTTACTCGCCAGTGCCTCAGGCT 2342
Db 562 AACAGCTGTAGAAGTGTGCAAGAAACAGGATAAGTTACTCGCCAGTGCCTCAGGCT 721
QY 2343 ATCAAGTTTCTGAAGCAAGATCAGCTGCTCTCTTCCACTCAACAATTATTCGCAATGA 2402
Db 722 ATTCAGTTTATGAAGAAACAGCTGCTGCTCTCTTCCACTCAGCGATTCGCGAGGA 781
QY 2403 CAGTCCCTCAGGACATGC 2421
Db 782 CATCTCCCTCAGGACATGC 800

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 817)

REFERENCE
AUTHORS Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utterbach, T., Feldblyum, T., Liang, F., Creasy, T. and Fraser, C.M.
TITLE Genomic survey sequencing of *Landisberg erecta* ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
JOURNAL Unpublished (2000)
COMMENT Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@tigr.org
For additional information, see <http://www.tigr.org/tdb/at/at.html>
Seq primer: TR
Class: shotgun.

FEATURES
Location/Qualifiers
1..817
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="LANDSBERG ERECTA"
/db_xref="taxon:3702"
/clone_lib="LERAW77"
/clone_lib="LERA"
/note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."

ORIGIN
Query Match 21.9%; Score 682.2; DB 28; Length 817;
Best Local Similarity 91.4%; Pred. No. 4e-114;
Matches 745; Conservative 0; Mismatches 68; Indels 2; Gaps 2;

QY 1312 AGGAGGAGCAATCGAGTTCTTCCACATCGAAGTGGAAATGTATGAGGACTCACAG 1371
DB 1 AGGAGGAGCAATCGAGTTCTTCCACGCTGCAAGTGGAGATGTATGAGGACTCACAG 60

QY 1372 ACTTTGAGCTGGATACCATCGGTTAGCTTCTTGTAGCTTCTTGTAGTGGATGAACAGCGGT 1431
DB 61 ACTTTGAGCTGGATACCATCGGTTAGCTTCTTGTAGCTTCTTGTAGTGGATGAACAGCGGT 120

QY 1432 ACCAGATTTAGATCAAGAAATTTGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAACCA 1491
DB 121 ATCAGCTTTTCATCAAGAGTTAGAGAACTGTTTGGTTTCCCTAGTGGAAAGGAACCA 179

QY 1492 ACCCAGGTTTGACAGGAGAGCTTAAGATTTGTGGCTTACTATTGGGAACATCTAC 1551
DB 180 AACCAAGATTTGACAGAGAGAGCTTAAGGACTTTGTGGGCACTATTGGGAACAGTCTAC 239

QY 1552 CGCTAAACTCGACGCGTCCAAGAGCAACCAATCCGAGTCCCTGTGATTCGCTACTTTC 1611
DB 240 CGTTAAACTCGGAGGTCACAGAGCAACCAATCCAGAGTCTTGTGATCGCTACTTTC 299

QY 1612 AGCCTCGGTTGCCAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTAACACAG 1671
DB 300 AGCAGTCGTTGCCAAGCTTTTACTCCAGGAGTCTACAGGACCGTGTCTAACACAG 359

QY 1672 ACATGAGATGATAGATCAGCGTTATAGGATTTCCGCTTACAAAGGAAGAGATG 1731
DB 360 ACATGAGATGATAGATGATGCGTTTACAGGATTTCTCCGCGGACGAGGAAGAGATG 419

QY 1732 TCCGTGAGGAGATCTTAACCACTCACCACAGTAATGCTTCTGTTGATCCATCTGTGTG 1791
DB 420 TCTTGAGCGGAGATCTCAAGACGACACCAAGTAATGCTTCTGTTGATCCATCTGTGTG 479

QY 1792 GGTACATGAAGTGGCGCTGCAAAACCGCAAGAAAGAGTGAAGAGGACATATGCGTGG 1851
DB 480 GATACAGGAAGTGGCGGTTGACGAACGGAAGAAAGTGAAGAGGACATATGCGTGG 539

QY 1852 GTGGGTTGTACGCCCAATCTGAAGTTTGTGGATTCGCTCAAGGAAGTAGGGTTAG 1911
DB 540 GTGGGTTGTGACGCCCAATTTTGCAGCTTTGTGGAGTACCCTCAACGAACGAGGTTAG 599

QY 1912 CACCGAGATGATGAGCTTGGATCACTTGCAGCGGATGTAGTCTTCTGAGTTGACATGG 1971
DB 600 CACCGAGATGATGAGCTTGGATCACTTGCAGCGGATGTAGTCTTCTGAGTTGACATGG 659

QY 1972 TTGGCGACTTTTCCCGCTACAGCTTTCGAGCTTTCATCGATTAGATCGCAACATCTTT 2031
DB 660 TTGGCGACTTTTCCCGCTACAGCTTTCGAGCTTTCATCGATTAGATCGCAACATCTTT 719

QY 2032 TCCCTCTGCAATTTACGCTACTAGGATTTCTGAGGCGAGCAACATTTGATTTCAAGCTTCGC 2091
DB 720 TCCCTCTGCAATTTACGCTACTAGGATTTCTGAGGCGAGCAACATTTGATTTCAAGCTTCGC 779

QY 2092 TTGAAGATCTTTTATTTTCGAGGCGAGCTCCGCCAACT 2126
DB 780 TTGAAGATCTTTTATTTTCGAGGCGAGCTCCGCCAACT 813

RESULT 5
BH244328 781 bp DNA linear GSS 13-NOV-2001
LOCUS AUJAA71TF AUJA Arabidopsis thaliana genomic clone AUJAA71, genomic survey sequence.
ACCESSION BH244328
VERSION BH244328.1 GI:16921164
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Utterback, T.V., Feldblyum, T.V. and Fraser, C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC f412
JOURNAL Unpublished (2001)
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1258.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..781
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="AUJAA71"
/note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into pHOS2 using BstXI linkers"

ORIGIN
Query Match 21.2%; Score 663.4; DB 28; Length 781;
Best Local Similarity 90.9%; Pred. No. 1.1e-110;
Matches 706; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 1322 AATCGAGTTCTTTCACACTCGAAGTGGAAATGTATGAGGACTCACAGACTTTGAGCT 1381
DB 5 AATCGAGTTCTTTCCTCCCTGCAAGTGGAGATGTATGAGGACTCACAGACTTTGAGCG 64

QY 1382 GGATACCATGGGTTAGGCTTCTTTCAGCTTCTTGTAGTGATGAACAGCGGTACCAAGTTTA 1441
DB 65 GGATACCATGGGTTAGGCTTCTTTCAGCTTCTTGTAGTGATGAACAGCGGTATCAGCTTC 124

QY 1442 GATCAAGAAATTTGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAACCAACCCAGGTT 1501
DB 125 GATCAAGAAATTTGAAGAACTGTTTGGTTTCCCTAGTGGAAAGGAACCAACCCAGGTT 184

QY 1502 TGACAGGAGAGAGCTTAAAGGATTTGTGGGCTACTATTGGGAACAATCTACCGCTAAACTC 1561


```
Db 185 TGACAGAGAAGAGCTTAAGGATTTGTGGGCCACTATAGGGAACAACTCTACCGTAAACTC 244
Qy 1562 GACGGCGTCCAAGACCAACAAATCCGGAGTCCCTGTGATTCGTACTTTTCAGCGCTCGGT 1621
Db 245 GCGCGGCTTAAGAGCAACCAATTCGAAGTCTCTGTATCCGCTACTTTTCAGCAATCGGT 304
Qy 1622 TGCCAAATGTTTTTACTCCAGGAGTCTACAGGCACCGTGTCTAACACAGACATGAAGAT 1681
Db 305 CGCCAAACGCTTTTACTCCAGGAGTCTACAGGCACCGTGTCTAACACAGACATGGAGAT 364
Qy 1682 GATAGATTCAGCGCTTATAGGATTCCTCGGCTTCAAAAAGGAAGATGTCCTGAGAGG 1741
Db 365 GATAGATTCAGCGCTTACAGGATTCCTCGCGGACGAAGGGAAGATGTCCTGAGAGG 424
Qy 1742 AGATCTTAAACGACTCACCAACAGTAATGCTCTGTGTATCATCTGTCTGGGTACATGAA 1801
Db 425 CGATCTCAACAACGACCAACAGTAATGCTCTGTGTATCATCTGTGTGATACAGGAA 484
Qy 1802 GTGGCGCTGACAAACGCAAGAAGAGTAAAGAGGACACTATGCGTGGGTGGCGTTGT 1861
Db 485 GTGGCGCTTGACGAACGGAAGAAGAGTGAAGAGGACACTATGCGTGGGTGGCGTTGT 544
Qy 1862 GACCCCAATCTGAAGTTTGTGGAGTCCGCTCAAGGAAGTATGAGTTCAGCACCGAGAT 1921
Db 545 GTCCCAATCTGAAGCTTGTGAGTACCGCTCAAGGAACGAGTTTATGACCGAGGAT 604
Qy 1922 GATGGACTTGGATCACTTGCCTCCGCTGAGTCTCTGAGTTTCACATGTTGCGGACTT 1981
Db 605 GATGGACTTGGATCACTTTCGCGAGTGTAGTCTTGTGAGTTTACATGTTGCGGACTT 664
Qy 1982 TCACCGCTACAGGTTTCGAGCAATTCATGATTAAGTTCGCAACATCTTTTCCCTCGAT 2041
Db 665 TCACCGCTACAGGTTTCGAGCAATTCATGATTAAGTTCGCAACATCTTTTCCCTCGAT 724
Qy 2042 TTACGCTACTAGGATTCGAGGCGAGCAATTCATGATTAAGTTCGCAACATCTTTTCCCTCGAT 2098
Db 725 TGACGCTACTAGGATTCGAGGCGAGCAATTCATGATTAAGTTCGCAACATCTTTTCCCTCGAT 781
```

```
RESULT 6
BH237552 848 bp DNA linear GSS 13-NOV-2001
LOCUS AU11B35TF AU11 Arabidopsis thaliana genomic clone AU11B35, genomic
DEFINITION survey sequence.
ACCESSION BH237552
VERSION BH237552.1 GI:16908149
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 848)
Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
Feildblum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC T2P24
Unpublished (2001)
Other GSSs: AU11B35TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 849.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .848
/organism="Arabidopsis thaliana"
/moi_type="genomic DNA"
/strain="Columbia"
```

FEATURES

source

```
/db_xref="taxon:3702"
/clone="AU11B35"
/clone_lib="AU11"
/note="Vector: PHOS2; Site 1: BstXI; 2-3 kb sheared BAC
DNA inserted into PHOS2 using BstXI linkers"
```

ORIGIN

```
Query Match 20.7%; Score 647.2; DB 28; Length 848;
Best Local Similarity 87.5%; Pred. No. 9.2e-108;
Matches 740; Conservative 0; Mismatches 83; Indels 23; Gaps 2;

Qy 2021 CAACATTCTTTTCCCTGCAATTTACGCTACTAGGATTCCTCGAGGCGAGCAACATTGACTT 2080
Db 5 CCACATTCTTCTCCCTGCAATTTACGCTACTCGGATTCCTCGAGGCGAGCAACATTGACTT 64
Qy 2081 CAAGCTTCGCGTTGAAGATCTTTTATTCGAGGGTAGTCCGCCAACTGAGGAGATCAGTCA 2140
Db 65 CAAGCTTCGCGTTGAAGATCTTTTATTCGAGGGTAGTCCGCCAACTGAGGAGATCAGTCA 124
Qy 2141 CACCGAAGGAGCTACAAATAGAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTT 2200
Db 125 TACCGAAGGAGCTACAAATAGAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTT 184
Qy 2201 TGACACGAGCATGATCATTTTTCAGTGAGCATATATACCTCCAGGAGGAAACGAGAGTTT 2260
Db 185 TGACACGAGCATGATCATTTTTCAGTGAGCATATATACCTCCAGGAGGAAACGAGAGTTT 244
Qy 2261 GAGCGAAGCTCACAGGAACAAACAAAGCTGCAAGTGTGCAAGAAACAGAGATAAGTT 2320
Db 245 GAGCGAAGCTCACAGGAACAAACAAAGCTGCAAGTGTGCAAGAAACAGAGATAAGTT 304
Qy 2321 ACTCGCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGCAAGATCAGTCTCTCTCTTC 2380
Db 305 ACTCGCAAGTGCCTCAGGGCTATCAAGTTTCTGAAGCAAGATCAGTCTCTCTCTTC 364
Qy 2381 CACTACAACTATTCCGCAATGACAGCTCCCTCAGGACATGCTTCGAGGAGATATGAGCG 2440
Db 365 CACTACAGGATTCGCGCAGGAGCATCTCCCTCAGGACATGCTTCGAGGAGATATGAGCG 424
Qy 2441 GCCCGA-----GCCTAGAGACAGAAATTCGATGTCCTCTGC 2479
Db 425 ACCCGACCAAGTCGACGCGAGGCGCTTTGCTTAGAGAGCAGGAGATTCGCGATGTCCTTC 484
Qy 2480 GAGGCAATTCATCATTCGAGGCTCGTGAATCTAGGAAGAATAGGAGACGACACTCACTCG 2539
Db 485 GAGGCAATTCATCATTCGAGGCTCGTGAATCTAGGAAGAATAGGAGACGACACTCACTCG 544
Qy 2540 ATCTAGCAGGAGGACGACGACTTCTGCACTCTCTGAGTTTACGCGACCGCGTGTGG 2599
Db 545 ATCTAGCAGTAGGAGCGCAGCTTCTGCACTCTCTGAGTTTACGCGACCGCGTGTGG 604
Qy 2600 CCGCAATAGGAAGAGAGAGTTCAGTATCTCAGAGCGGTGCTGGCGCGCACAGAGCTGA 2659
Db 605 CCGCAATAGGAAGAGAGAGTTCAGTATCTCAGAGCGGTGCTGGCGCGCACAGAGCTGA 664
Qy 2660 TGAGATCGAGTACCCACATCTGAGCTGATACGGAACATGGCGGTTTCGCTATGGCTTG 2719
Db 665 TGAGATCGAGTACCCACATCTGAGCTGATACGGAACATGGCGGTTTCGCTATGGCTTG 724
Qy 2720 GGAGCAATCACAGGCGCCATTCGACTACCAACTTCGTTTCAATTCGACTGAGGTAAAGCG 2779
Db 725 GGAAGCATCGAGCGAGCCATTCGAGGCACTACGTTTCAATTCGACTGAGGTAAAGCG 784
Qy 2780 CCTCACTTCAACATTATATATATCTCTGATGATGTTTCTTTATTTTGTGTTTGTGTTG 2839
Db 785 CCTCACTTCAACAT--GGATATATATCTCTTGGGATGCGGTTCTCATTTTTTGTGTTG 842
Qy 2840 ATTGGA 2845
Db 843 GATGGA 848
```

RESULT 7

```
BH244305/c
LOCUS BH244305 813 bp DNA linear GSS 13-NOV-2001
DEFINITION AUIVA072TF AUIV Arabidopsis thaliana genomic clone AUIVA072, genomic survey sequence.
ACCESSION BH244305
VERSION BH244305.1 GI:16921118
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 813)
AUTHORS Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Uterback, T.V., Feldblyum, T.V. and Fraser, C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC f4i2
JOURNAL Unpublished (2001)
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1258.
Seq primer: TF
Class: sheared ends.
FEATURES             Location/Qualifiers
     source            1..813
                        /organism="Arabidopsis thaliana"
                        /mol_type="genomic DNA"
                        /strain="Columbia"
                        /db_xref="taxon:3702"
                        /clone="AUIVA072"
                        /clone_lib="AUIVA"
                        /notes="Vector: pPOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into pPOS2 using BstXI linkers"
ORIGIN
Query Match          19.9%; Score 622.6; DB 28; Length 813;
Best Local Similarity 87.5%; Pred. No. 2.8e-103;
Matches 705; Conservative 0; Mismatches 69; Indels 31; Gaps 1;
QY 978 GTGAGGACACATGTTTACGAGAGCTATAGGGATGAATTTGAACGCTCTGCAGCTCGAGC 1037
DB 805 GAGAGAACACATGTTTACCAGAGCTTCAGGGATGAATTCGAACGCTCTGCAGCCGCGAGC 746
QY 1038 TAATCAAGAGAGCTGAATCGTAGAGAAAGAGGGCGCATCTCGAGTAGATAGAGCT 1097
DB 745 TAATCAGAGAAGGGCTGAATTCGCCAGAGAAAGAGGGCGCATGTCAAGTAGGTAGGATGAGCT 686
QY 1098 GATTGATGAGGATCAAACTGAGTAGAGCCAGAGTCAATGGCGCAAGGAGACGAAGCT 1157
DB 685 GATTGATGAGGCTATCGAAGTAGATGATGAGCCGAGTCAATGGCGCAAGGAGACGAAGCT 626
QY 1158 ACTGAACAAATCCGACGAGGTTCAGTGGAGGAGTATATCAGATCTTTTTCAGATGAATGA 1217
DB 625 GCTGAACAAACACGACGAGGTTCAGTGGAGGAGTATATCAGGTCTTTTCAGATGAATGA 566
QY 1218 CTTCTGGGGAACGAGGTATCCCTGATATGAGACTTTAGCCCGAGTTGGGGTTACTGGAGGA 1277
DB 565 TTTCTGGGGAACGCGGTATCCGTCATGAGACTTTAGCCCGAGTTAGGGTTACTGGAGGA 506
QY 1278 CGTGCACATCTGTTTCGAGAAGTGTCAATCTG----- 1308
DB 505 CGTGCACATCTGTTTCGAGAAGTGTCAATCTGGAGAACTCATGTCCTTACCCCTTACCCTGC 446
QY 1309 --ATAAGGAGAGACAATCGAGTTCTTTTCACACTGCAAGTGGAATGATGAGGAGCT 1366
DB 445 CTATAGGAGAGACAATCGAGTTCTTTTCACACTGCAAGTGGAATGATGAGGAGCT 386
QY 1367 CACAGACTTTGAGTGCATACCGGGTTCAGGCTTCTTCAGCTTCTTAGTGGATGAACA 1426
DB 385 CACGGACTTTGAGCGGGATACCATGGGGTTAGGATTTCTTGACGTTCTCAGTGAATGAACA 326
```

```
1427 GCGGTACCAGATTAGATCAAGAAATGGAAGAACTGTTTGGTTTCCCTAGTGAAGGG 1486
      |||||
325 GCGGTATCAGCTTTCGATCAAGAACTGGAAGAACTGTTTGGTTTCCCTAGTGAAGGG 266
      |||||
1487 AACCAACCCAGGTTTGACAGGAGAGCTTAAGGATTGTGGGCTACTATTGGGAACAA 1546
      |||||
265 AACCAACCCAGGATTGACAGGAGAGCTTAAGGATTGTGGGCTACTATTGGGAACAA 206
      |||||
1547 TCTACCGCTAAACTCGACGCGGTCCAAGACCAACCAATCCGAGTCTCTGTGATTGCTA 1606
      |||||
205 TCTACCGCTAAACTCGGCGGCTTAAGAGCAACCAATTCGAAGTCCGTGTGATCGCTA 146
      |||||
1607 CTTTCAGGCTCGGTTGCAATGTTTTTACTCCAGGAGTCTACAGGACCGTGTCTAA 1666
      |||||
145 CTTTCAGCAATCCGTCGCCAAGCTCTTTTACTCCAGGAGTCTACAGGACCGTGTCTAA 86
      |||||
1667 CACAGACATGAGATGATAGATTACGCGCTTATAGGATTCCTCCGCTTACAAAGGAAA 1726
      |||||
85 CACGACATGAGATGATAGATTACGCGCTTATAGGATTCCTCCGCTTACAAAGGAAA 1726
      |||||
1727 GAATGCTCTGAGAGGAGATCTTAAC 1751
      |||||
25 GAATGCTCTGAGAGGCGATCTCAAC 1
      |||||
RESULT 8
LOCUS BH235546 806 bp DNA linear GSS 13-NOV-2001
DEFINITION AUIVA072TF AUIV Arabidopsis thaliana genomic clone AUIVA07, genomic survey sequence.
ACCESSION BH235546
VERSION BH235546.1 GI:16905904
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 806)
AUTHORS Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Uterback, T.V., Feldblyum, T.V. and Fraser, C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC F5K16
JOURNAL Unpublished (2001)
COMMENT Other GSSs: AUIVA07TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1269.
Seq primer: TF
Class: sheared ends.
FEATURES             Location/Qualifiers
     source            1..806
                        /organism="Arabidopsis thaliana"
                        /mol_type="genomic DNA"
                        /strain="Columbia"
                        /db_xref="taxon:3702"
                        /clone="AUIVA07"
                        /clone_lib="AUIV"
                        /notes="Vector: pPOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into pPOS2 using BstXI linkers"
ORIGIN
Query Match          19.6%; Score 611.6; DB 28; Length 806;
Best Local Similarity 87.5%; Pred. No. 2.8e-101;
Matches 705; Conservative 0; Mismatches 69; Indels 32; Gaps 2;
QY 1182 AGTGGAGAGTATATCAGATCTTTTTCAGATGAATGATCTTCGGGGAACGAGGTATCCCTG 1241
DB 1 AGTGGAGAGTATATCAGATCTTTTTCAGATGAATGATCTTCGGGGAACGAGGTATCCCTG 60
```

QY	1242	ATATGAGACTTTAGCCCACTGGGGTTACTGAGGACCGTCAGACATCTGTTTCGAGAAAGTG	1301
Db	61	CTATGAGACTTTAGCCCACTAGGGTTACTTGAGGACCGTCAGACATCTGTTTCGAGAAAGTG	120
QY	1302	TCATCTG-----ATAGGAGGAGACAATCGAGTT	1330
Db	121	TCATCTGGAGAAACTCATGTCTTACCCTTACCTCTATTAAGGAGGAGACAATCGAGTT	180
QY	1331	TCCTTTCCACACTGCAAGTGGAAATGATGAGGACTCACAGACTTCGAGCTGGATACCAT	1390
Db	181	TCCTTTCCACGCTGCAAGTGGAGATGATGAGGACTCAGCGACTTCGAGCGGGATACCAT	240
QY	1391	GGGGTTAGGCTCTTTGACGTTCTTTAGTGGATGAACAGCGGTACCAGATTTTAGATCAAGAA	1450
Db	241	GGGGTTAGGATTTCTTGACGTTCTCAGTGAATGAACAGCGGTATCAGCTTTTCGATCAAGAA	300
QY	1451	ATTGGAACAACGTTTGTGGTTTCCCTAGTGGAAAGGGAACCAACCCAGGTTTCACAGGGA	1510
Db	301	GTTGGAGAACTGTTTGGTTTCCCTAGTGGAAAGGGAACCAACCCAGATTTTCACAGAGA	360
QY	1511	AGAGCTTAAGGATTTGTGGGCTACTATTGGGAACAATCTACCGCTAACTTCGAGCGGGTC	1570
Db	361	AGAGCTTAAGGATTTGTGGGCCACTATAGGGACAATCTACCGCTAACTTCGAGCGGGTC	420
QY	1571	CAAGAGCAACCAATCCGAGTCTGTGATTCGCTACTTTCAGCGGTCGGTTCGCCAATGT	1630
Db	421	TAAGAGCAACCAAAATTCGAATCTGTGATTCGCTACTTTTCAGCAATCGGTTCGCCAATGT	480
QY	1631	TTTTTACTCCAGGAGGCTACAGGACCGGTCTTACACAGACATGAAGATGATGATTC	1690
Db	481	CTTTTACTCCAGGAGGCTACAGGACCGGTCTTAAACGGAATGAGATGATGATTA	540
QY	1691	AGCGCTTATAGGGATTCGCGCCTTACAAAGAAAGAAATGCTCTCAGAGGAGATCTTAA	1750
Db	541	TGCGCTTACAGGATTCCTCGCGGACGAAGGGAAGATGTTTTCAGAGGGCGATCTCAA	600
QY	1751	CGACTCACACACAGTAAATGCTCTGTGTATCCATCTGTGTGGGTAATGAAGTGGGCGCT	1810
Db	601	CAACGCACCAACCAAGTAAATGCTCTGTGTATCCATCTGTGTGGATACAGAAATGGGCGTT	660
QY	1811	GACAAACGGCAAGAAAGGTAGAGGAGCACTATCGGTGGGTGGGCTTGTCAACCCCAAT	1870
Db	661	GACGAACGGGAAGAAAGGTAGAGGAGCACTATCGGTGGGTGGGCTTGTCTGCCCAAT	720
QY	1871	TCTGAAAGTTTGTGGAGTTCCGCTCAAGGAAGTAGGGTTAGCACCGAGATCATCGACTT	1930
Db	721	TCTGGAAGCTTGTGGAGTACCGCTCAAGG-ACCAGGTTTAGCACCAGGATGATGACTT	779
QY	1931	GGATCACTTTCGCCGAGATGTGAGTTCT	1956
Db	780	GGATCACTTTCGCCAGATGTGAGTTCT	805

RESULT	9
LOCUS	BH244756/c
DEFINITION	linear GSS 13-NOV-2001
ACCESSION	AUIJAO07F AUIJ Arabidopsis thaliana genomic clone AUIJA07, genomic survey sequence.
VERSION	BH244756
KEYWORDS	BH244756.1 GI:16921934
SOURCE	GSS.
ORGANISM	Arabidopsis thaliana (thale cress)
	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE	1 (bases 1 to 681)
AUTHORS	Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V., Fel'dlyum,T.V. and Fraser,C.M.
TITLE	Survey sequencing of Arabidopsis thaliana BAC TI9J2
JOURNAL	Unpublished (2001)
COMMENT	Other GSSs: AUIJA07TR

Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 975.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .681
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AU1JA07"
/clone_lib="AU1J"
/notes="Vector: pPOS2; Site 1: BstXI; 2-3 Kb sheared BAC
DNA inserted into pPOS2 using BstXI linkers"

ORIGIN

Query Match	19.4%;	Score 604.6;	DB 28;	Length 681;
Best Local Similarity	93.5%;	Pred. No. 5.7e-100;		
Matches 631;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;
QY	2000	GCATTCATCGATTAGTAATCGCCAAACAATCTTTTCCCTGCATTTACGTACTACGATTCT	2059	
Db	681	GCATTCATCGATTAGTAATCGCCAAACAATCTTCTCCCTTACATTGACGCTACAAGGATTC	622	
QY	2060	CGAGGCGACGAAACATTGACTTCAAGCTCGCGTTGAAGATCTTTATTTCGAGGGCAGTCC	2119	
Db	621	CGAGGCGACGAAACATTAACTTCAGGCTCGCGTTGAAGATCTTTATTTCGAGGGCAGTCC	562	
QY	2120	GCCAACTGAGGAGATTAGTCAACCCGAAGGAGCTCAATAGAAAGATGTTGATGAGACATA	2173	
Db	561	GCCAACTGAGGAGATCAGTCAACCCGAAGGAGCTACAACAGAAGATGTTGATGAGACAGA	502	
QY	2180	TCATATAGATCAGGCGGAGTTTGACACAGAGCATGTATCATTTTCAGTGAGCATATACCTCC	2239	
Db	501	TGATATAGATCAGGCGGAGCTTTGACACAGAGCATGTATCATTTTAAGTGAGCATATACCTCC	442	
QY	2240	AGCGAGGAAAAAGCAAGAGCTTTGAGCGAAGCTCACAGGAACACAGCAGCTGCAGAAGTG	2299	
Db	441	AACGAGGGAAGGAAGAGCTTTGAGCGAAGCTCACAGGAACACAGCAGCTGTAGAAGTG	382	
QY	2300	GTGCAAGAAAACAGGATAAGTTACTCGCAAGTGCTCAGGGCTATCAAGTTCTTGAAGGA	2359	
Db	381	GTGCAAGAAAACAGGATAAGTTACTCGCAAGTGCTCAGGGCTATCAAGTTCTTGAAGGA	322	
QY	2360	CAAGATCAGCTGCTCCTTCCACTCAACAATTTCGCGAATGACAGTCCCTCAGGACAT	2419	
Db	321	CAAGATCAGCTGCTCCTTCCACTCAGCTATTTCCGACGGACAGCTCCCTCAGGACAT	262	
QY	2420	GCCTTCGAGGAGATATGACGCGCCGCGAGCCTTAGAGACAGAAGATTCTGCATGTCCTCG	2479	
Db	261	GCCTTCGAGGAGATATGACGTCCTGAGCCTAGAGAGTAGAGATTCGATGTTCCCGC	202	
QY	2480	GAGCATTATCATTCGAGCCTCGTGAATCTAGGAAGATTAGGAGAACGACATCTCACTCG	2539	
Db	201	AAGGCATTTCATTCATTGAGCCTCATGAATCTAGGAGGAAGAGGACACACTCACTCG	142	
QY	2540	ATCTAGCAGCAGGACAGACGACTTCGAGCTCTGTAGTTTACGCGACCGCGGTGCTGG	2599	
Db	141	ATCTAGCAGCAGGACAGACGACTTCTGCACTCTGTACTTTTACGCGACCGCGGTGCTGG	82	
QY	2600	CCGCAATAGAAAGAGAGAGGTTCAGTATCCTCAGACGGGTGCTGGCCGCCACAGAGCTGA	2659	
Db	81	CCGCAAGCAGGAGAGAGGTTCAGTATCCTCAGACGGGTGCTGGCCGCCACAACTGA	22	
QY	2660	TGAGATCGAGTACCC	2674	
Db	21	TGAGGTCGAATACCC	7	

2572 CTCTAGTTTACGGACCGCGGTCTGCGCCGCAATAGAGAGAGAGGTTCGAGTATCTCT 2631
 421 CTCTACTTTACGGACCGCGGTCTGCGCCGCAATAGAGAGAGAGGTTCGAGTATCTCT 480
 2632 AGAGCGGTCTGCGCCGCAATAGAGAGAGGTTCGAGTATCTCTCTGAGTATCTCT 2691
 481 AGAGCGGTCTGCGCCGCAATAGAGAGAGGTTCGAGTATCTCTCTGAGTATCTCT 540
 2692 CGGACATGCGCGGTCTGCGCCGCAATAGAGAGAGGTTCGAGTATCTCTCTGAGTATCTCT 2751
 541 CAGACAAAGCGGTCTGCGCCGCAATAGAGAGAGGTTCGAGTATCTCTCTGAGTATCTCT 600
 2752 TTTGTTTATTTTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCT 2811
 601 TAGCTTCTCTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCT 660
 2812 GTGATTTCTCTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCT 2832
 661 GTGATTTCTCTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCTCTGAGTATCTCT 681

RESULT 11
 B08362 1101 bp DNA linear GSS 14-MAY-1997
 LOCUS T24116-T7.1 TAMU Arabidopsis thaliana genomic clone T24116, genomic
 DEFINITION survey sequence.
 ACCESSION B08362
 VERSION B08362.1 GI:2089483
 KEYWORDS GSS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
 Ecker, J.
 TITLE BAC End Sequences at ATGC
 JOURNAL Unpublished (1997)
 COMMENT Other GSSs: T24116-T7, T24116-Sp6, T24116-Sp6.1
 Contact: Ecker J.
 Arabidopsis Thaliana Genome Center
 University of Pennsylvania
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA
 19104
 Tel: 215-898-9384
 Fax: 215-898-8780
 Email: jecker@atgenome.bio.upenn.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 133
 High quality sequence stop: 810.
 FEATURES
 Location/Qualifiers
 1..1101
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T24116"
 /sex="hermaphrodite"
 /clone_lib="TAMU"
 /note="Vector: BelobACII; Site 1: HindIII; Site 2:
 HindIII; Produced by Rod Wing"

Query Match 19.2%; Score 598.6; DB 28; Length 1101;
 Best Local Similarity 81.9%; Pred. No. 5.8e-99;
 Matches 803; Conservative 0; Mismatches 147; Indels 31; Gaps 9;
 1873 TGAAGTTTGTGAGTTCGCTCAAGCAAGTGGTTAGCAGCAATGATGACTTGG 1932
 84 TGCTTCTGTTGAGTTCCTCGAGAACCCAGGTTTGGACCCCGGATGATGACTTGG 143

RESULT 10
 BH244796 691 bp DNA linear GSS 13-NOV-2001
 LOCUS AUIJAS58T AUIJ Arabidopsis thaliana genomic clone AUIJAS8, genomic
 DEFINITION survey sequence.
 ACCESSION BH244796
 VERSION BH244796.1 GI:16921996
 KEYWORDS GSS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 691)
 AUTHORS Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Utterback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.
 TITLE Survey sequencing of Arabidopsis thaliana BAC TIJ92
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: AUIJAS8TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 Prom Wash. U contig 975.
 Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers
 1..691
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="AUIJAS8"
 /clone_lib="AUIJ"
 /note="Vector: pPOS2; Site 1: BstXI; 2-3 kb sheared BAC
 DNA inserted into pPOS2 using BstXI linkers"

Query Match 19.3%; Score 601; DB 28; Length 691;
 Best Local Similarity 92.7%; Pred. No. 2.6e-99;
 Matches 631; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 2152 CTACAAATAGAGATGTTGATGAGACATATGATAGATGAGCGGAGTTTGACAGCA 2211
 1 CTACACAGAGATGTTGATGAGACATATGATAGATGAGCGGAGTTTGACAGCA 60
 2212 TGTATCATTTGATGAGATATACCTCCAGCGAGAAAGCAAGAGTTTGACGAGCTC 2271
 61 TGTATCATTTAAGTGAGCATATACCTCCAGCGAGAAAGCAAGAGTTTGACGAGCTC 120
 2272 ACAGCAACACAGCAAGCTGCAAGAGTGTGCAAGAAACAGAGTAAGTTTCTCGCAAGT 2331
 121 ACAGCAACACAGCAAGCTGCAAGAGTGTGCAAGAAACAGAGTAAGTTTCTCGCAAGT 180
 2332 GCTCAGGCTATCAAGTTTCTGAAGCAACAGATGAGTGTCTTCCACTACACTA 2391
 181 GCTCAGGCTATCAAGTTTCTGAAGCAACAGATGAGTGTCTTCCACTACACTA 240
 2392 TTCCGCAATGACATGCTCCTCAGGACATGCTTCCAGGAGATATGAGTGTCTTCCACTA 2451
 241 TTCCGCAATGACATGCTCCTCAGGACATGCTTCCAGGAGATATGAGTGTCTTCCACTA 300
 2452 GAGAGCAGAGATTTCTGCATGTCCCTCGGAGGATTCATCATTCGAGCTGTGGAATCTA 2511
 301 GAGAGTGGAGATTTCTGCATGTCCCTCGGAGGATTCATCATTCGAGCTGTGGAATCTA 360
 2512 GGAAGATAGGAAACACACTCACTGATCTAGACAGGAGCAGACGCTTCTCAGT 2571
 361 GGAGGAGAGAGAGGACACACTCACTGATCTAGACAGGAGCAGACGCTTCTCAGT 420

Db 165 AGCATTCATCGAATTAATTCGCAACATCTCTTCCCTGCAATTAACGCTACTAGATTC 106

QY 2059 TCGAGGCGAGGAACATTCAGTCTCAAGCCCTCGCCTTGAAGATCTTTATTTTCGAGGCGAGTC 2118

Db 105 TCGAGGCGAGGAACATTCAGTCTCAAGCCCTCGCCTTGAAGATCTTTATTTTCGAGGCTAGTC 46

QY 2119 CCGCAACTGAGGAGATTAATTCACCGAAGGAGCTACAATAGAAG 2163

Db 45 CGCCAACTGAGGAGATCACTCATAGAAAACGAGCTACAACAGAAG 1

RESULT 13

BH237151/c

LOCUS

DEFINITION

792 bp DNA linear GSS 13-NOV-2001

AUII06TR AUII Arabidopsis thaliana genomic clone AUII06, genomic survey sequence.

ACCESSION

BH237151

GI:16907522

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 792)

Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Utterback, T.V., Feldblyum, T.V. and Fraser, C.M.

Survey sequencing of Arabidopsis thaliana BAC T2P24

Unpublished (2001)

Other_GSSs: AUII06TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

From Wash. U contig 849.

Seq primer: TR

Class: sheared ends.

FEATURES

source

1..792

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="AUII06"

/clone_lib="AUII"

/note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into pHOS2 using BstXI linkers"

ORIGIN

Query Match 18.8%; Score 585.6; DB 28; Length 792;

Best Local Similarity 87.1%; Pred. No. 1.5e-96;

Matches 679; Conservative 0; Mismatches 69; Indels 32; Gaps 2;

QY 1223 GGGGAACGAGTATCCCTGATATGAGACTTTAGCCACTTGGGTTACTGGAGGACGTGC 1282

Db 779 GGGGAACGAGTATCCCTGATGAGACTTTAGCCACTTGGGTTACTGGAGGACGTGC 720

QY 1283 AGCATCTCTCGAAGTGTCTATCTG-----ATA 1311

Db 719 AGCATCTCTCGAAGTGTCTATCTGGAGACACTCATGTCTTACCTTTACCTGCTGCTATA 660

QY 1312 AGGAGGACAAATCAGTCTTTCTTCCACACTGCAAGTGAATGTATGAGGACTCACAG 1371

Db 659 AGGAGGACAAATCAGTCTTTCTTCCAGCTGCAAGTGAATGTATGAGGACTCACAG 600

QY 1372 ACTTTGAGTGGATACCATCGGTTAGGCTTCTTCCAGCTTCTTGTGGATGAACAGCGGT 1431

Db 599 ACTTTGAGTGGATACCATCGGTTAGGCTTCTTCCAGCTTCTTGTGGATGAACAGCGGT 540

QY 1432 ACCAGATTTAGATCAGAAATTTGGAGAACTCTTTGGTTTCCTAGTGGAAAGGGAACCA 1491

Db 539 ATCAGCTTCGATCAAGAAGTTGGAAAGCTGTTTGGTTTCCCTTGTGGAAAGGGAACCA 480

QY 1492 ACCCAGGTTTGACAGGAAGAGCTTAAGGATTTTGTGGGCTACTATTGGGAACAATCTAC 1551

Db 479 AACCCAGATTTTGACAGGAAGAGCTTAAGGATTTTGTGGGCTACTATTGGGAACAATCTAC 420

QY 1552 CCCTAAACTCGACGCGGTCACAGCAACCAATCCGAGTCTCTGTGATCGCTACTTTTC 1611

Db 419 CGCTAACTCGGCGGGTTTAAGGCAACCAATCCGAGTCTCTGTGATCTGTACTTTTC 360

QY 1612 AGCGCTCGGTTCCCAATGTTTTTACTCCAGGAGCTTACAGSCCGGTCTTAACACAG 1671

Db 359 AGCGATCGGTCGCCAACGCTTTTACTCCAGGAGCTTACAGACCCGTTCTTACACCG 300

QY 1672 ACATGAAGATGATAGATTCAGCGCTTATAGGAGTCTCCGCCCTTTACAAAGGAAGAATG 1731

Db 299 ACATGGAGATGATAGATTTATGCTTACAGGATTTCTCCGCCGACGAAGGGAAGAATG 240

QY 1732 TCCTGAGAGGAGATCTTAACGACTCACACCGATGATGCTCTGTGATCCATCTCTGTG 1791

Db 239 TCCTGAGAGGAGATCTCAACCAACGACCACTTATGCTCTGTGATCCATCTCTGTG 180

QY 1792 GGTACATGAGTGGCGCTGCAAAACGCAAGGAAGGTAAAGAGGACACTATGCGTGG 1851

Db 179 GATACAGGAAGTGGCGTTGACGACGGAAGGAAGGTGAGAGGACACTATGCGTGG 120

QY 1852 GTGCGCTTGTGACGCGCAATTTCTGAAAGTTTGTGGAGTTCGCTCAAGGAAGTAGGTTAG 1911

Db 119 GTGCGCTTGTGACGCGCAATTTCTGAAAGTTCGAGTTCGCTCAAGGAAGTAGGTTAG 60

QY 1912 CACCGAGATGATGAGTTCGATCCTTGGCGCGATGAGTTCCTCTGAGTTTGACATGG 1971

Db 59 CACCGAGGATGATGAGTTCGATCCTTGGCGCGAT-AGAGTTCTTGGAGTTTGACATGG 1

RESULT 14

BH237213

LOCUS

DEFINITION

670 bp DNA linear GSS 13-NOV-2001

AUII028TR AUII Arabidopsis thaliana genomic clone AUII028, genomic survey sequence.

ACCESSION

BH237213

GI:16907584

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 670)

Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Utterback, T.V., Feldblyum, T.V. and Fraser, C.M.

Survey sequencing of Arabidopsis thaliana BAC T2P24

Unpublished (2001)

Other_GSSs: AUII028TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

From Wash. U contig 849.

Seq primer: TR

Class: sheared ends.

FEATURES

source

1..670

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="AUII028"

/clone_lib="AUII"

/note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into pHOS2 using BstXI linkers"

ORIGIN

Query Match 18.8%; Score 585.6; DB 28; Length 792;

Best Local Similarity 87.1%; Pred. No. 1.5e-96;

Matches 679; Conservative 0; Mismatches 69; Indels 32; Gaps 2;

QY 1223 GGGGAACGAGTATCCCTGATATGAGACTTTAGCCACTTGGGTTACTGGAGGACGTGC 1282

Db 779 GGGGAACGAGTATCCCTGATGAGACTTTAGCCACTTGGGTTACTGGAGGACGTGC 720

QY 1283 AGCATCTCTCGAAGTGTCTATCTG-----ATA 1311

Db 719 AGCATCTCTCGAAGTGTCTATCTGGAGACACTCATGTCTTACCTTTACCTGCTGCTATA 660

QY 1312 AGGAGGACAAATCAGTCTTTCTTCCACACTGCAAGTGAATGTATGAGGACTCACAG 1371

Db 659 AGGAGGACAAATCAGTCTTTCTTCCAGCTGCAAGTGAATGTATGAGGACTCACAG 600

QY 1372 ACTTTGAGTGGATACCATCGGTTAGGCTTCTTCCAGCTTCTTGTGGATGAACAGCGGT 1431

Db 599 ACTTTGAGTGGATACCATCGGTTAGGCTTCTTCCAGCTTCTTGTGGATGAACAGCGGT 540

QY 1432 ACCAGATTTAGATCAGAAATTTGGAGAACTCTTTGGTTTCCTAGTGGAAAGGGAACCA 1491

DNA inserted into pHOS2 using BstXI linkers"

ORIGIN	Query Match Best Local Similarity Matches 612; Conservative 0; Mismatches 58; Indels 0; Gaps 0;	DNA inserted into pHO2 using BstXI linkers*
QY	1612	AGCGCTCGGTTGCCAANTGTTTTTACTCCAGGAGTCTACAGGCACCGTGCTTAACACAG 1671
Db		
QY	1	AGCGATCGGTGCCAACGCTTTTACTCCAGGAGTCTACAGACACCGTGCTTAACACGG 60
Db		
QY	1672	ACATCAGATGATGATTCACGGCTTATAGGATCTCCGCGCTTACAAAGGAAAGAAATG 1731
Db		
QY	61	ACATGGAGATGATGATTTATCGCTTTACAGGGATTCTCCGCGACGAAAGGGAAGAAATG 120
Db		
QY	1732	TCCTGAGAGGAGATCTTAAACGACTCACACACAGTAATGCCTCTGTTCATCCATCTGTGTG 1791
Db		
QY	121	TCTTGAGGGGATCTCAACACGCCACCAAGTAATGCCTCTCTGTATCCATCTGTGTG 180
Db		
QY	1792	GGTACATGAAGTGGCGCTGCACAAACGGCAGAAAGGTAAGAGGAGCACTATGCCTGG 1851
Db		
QY	181	GATACAGGAAGTGGCGCTTGACGAACGGGAAGAAAGGTGAGAGGAGCACTATGCCTGG 240
Db		
QY	1852	GTGCGGTGTGACGCCCAATTCGNAAGTTTGTGAGTTCGCTCAAGGAAGTAGGGTTAG 1911
Db		
QY	241	GTGCGGTATGACGCCCAATTCGGAAGCTTGTGAAGTACCGCTCAAGGAACCTAGGTTTAC 300
Db		
QY	1912	CACCGAATGATGGACTTGCGATCCTTGCGCCGATGTGAGTTCTCTGAGTTTGACATGG 1971
Db		
QY	301	CACCGAGGATGATGGACTTGGAATCATTTGCGCAGATGTGAGTTCTTGAGATTTGACATGG 360
Db		
QY	1972	TTGGCGACTTTCACCGCTACAGGTTGAGCAATTCATCGATTAGAATGCGCAACATCTTT 2031
Db		
QY	361	CTGGTGACTTTCACCGCTACAGGTTGAGCAATTCATCGATTAAAAATCGCCCAACATCTTC 420
Db		
QY	2032	TCCCTCTGATTTACGCTACTAGGATTCGAGGGCAGGAACATGTGATTCAGACCTGCGC 2091
Db		
QY	421	TCCCTCTGATTAACGCTACTAGGATTCGAGGGCAGGAACATGTGATCTCAAGCCTGCGC 480
Db		
QY	2092	TTGAAGATCTTTATTTTCGAGGGCAGTCCGCCAACTGAGGAGATTAGTTCACCCGAAGGAG 2151
Db		
QY	481	TTGAAGAACTTTATTTTCGAGGGTAGTCCGCCAACTGAGGAGTCACTATACCGAACGAG 540
Db		
QY	2152	CTACAATAGAAGATGTTGATGAGACATATGATATAGATGAGGCGGAGTTTGACACGAGCA 2211
Db		
QY	541	CTACAAACAGAAGATATTGATGAGACGGATGATATAGATGCGACGGAGTTTGACACGAGCA 600
Db		
QY	2212	TGTATCATTTCACTGAGCATATACCTTCGAGGAGGAAAGCAGAGCTTTGAGCGAGCTC 2271
Db		
QY	601	TGTATCATTTCACTGAGCATATACCTTCGAGGAGGAAAGCAGAGCTTTGAGCGAGCTC 660
Db		
QY	2272	ACAGGAACAA 2281
Db		
QY	661	ACAGGAACAA 670
Db		

RESULT 15	
BH244293	
LOCUS	
DEFINITION	756 bp DNA linear GSS 13-NOV-2001
ACCESSION	AUJAA06TF AUJA Arabidopsis thaliana genomic clone AUJAA06, genomic
VERSION	survey sequence.
KEYWORDS	
SOURCE	BH244293.1 GI:16921094
ORGANISM	GSS.
	Arabidopsis thaliana (thale cress)
	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
REFERENCE	1 (bases 1 to 756)
AUTHORS	Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V.,
	Feldbylum,T.V. and Fraser,C.M.
TITLE	Survey sequencing of Arabidopsis thaliana BAC f4i2

JOURNAL
COMMENT

Unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1258.
Seq primer: TF
Class: sheared ends.

FEATURES
source

ORIGIN

Query Match	18.5%;	Score 576.2;	DB 28;	Length 756;
Best Local Similarity	88.5%;	Pred. No. 8.1e-95;		
Matches 668;	Conservative 0;	Mismatches 63;	Indels 24;	Gaps 3;
QY	2237	TCCAGCGAGGAAAGCAAGAGTTTGGACGAAAGCTCACAGAAACAAACAGCAAGCTGCAGAA	2296	
Db	2	TCTTCGAGGGGAGCAGGAGTTTCGACGAAAGCTAAACAGAAACAAACAGCAAGCTGCAGAA	61	
QY	2297	GTGGTGCAAGAAACAGGATAAGTTACTCGGCCAAGTGCTCAGGGCTATCAAGTTTCTGAA	2356	
Db	62	GTGGTGCAATACACAGGACAGTTACTCGGCCAAGTGCTCAGGGCTATCAAGTTTATGAA	121	
QY	2357	GGACAGATCAGCTGCTCTCTTTCCACTAAACTATTCGCGCAATGACAGCTCCCTCAGGA	2416	
Db	122	GAACAAGCTCAGCTGCTCTCTTTCCACTATAGTTATTCGCGAGGACATCTCCCTCAGGA	181	
QY	2417	CATGCTTCGAGGAGATACGCGGCC-----GAGCCTAGAGA	2455	
Db	182	CATGCTTTCAGAGAGATACGCGGCCGAGCCAGTCGGCGCAGGCCCTGAGCCTAGAGA	241	
QY	2456	GCAGAAAGATTCTGCATGTCCTTCGCGAGGCATTCATCATTCGAGCCTCGTGAATCTAGGAA	2515	
Db	242	GCAGGTGATTCCGATGTCCTTACGAGGCATTCATCATTCGAGCCTCGTGAATCTGGAG	301	
QY	2516	GAATAGGAGAACGACACTCACTCGATCTAGCAGAGGAGCAGACGACTTCTCGAGTCTCG	2575	
Db	302	GAAGAGGAGGACACACTCTCGATCTAGCAGTAGGAGCGGACGACTTCTCGAGTCTCG	361	
QY	2576	TAGTTTACCGCACCGCGGTCTGGCGCAATAGAAAGAGAGGTCGAGTATCCTCAGAG	2635	
Db	362	TAGTTTACCGCACCGCGGTCTGGCGCAGCAAGAGGAGGTCGAGTTTCTCTCAGAG	421	
QY	2636	CGGTGTCGCGCCACAGAGCTGATGAGATCGAGTACCCACATGCTGGAGCTGATACGGA	2695	
Db	422	CGGTGTTGGCCGCCACAGAGCTGATGAGTTCGAGTACCCACCCTGCTGGAGCTGATACAGA	481	
QY	2696	ACATCGCGTTCGTCTATGGCTTGGGAGCAATCACAGGAGCCATTGACTACCAACTTCG	2755	
Db	482	ACAAGCGAGTTCGTCTATGGCTGGGAGCAATTCGAGGAGCCATTGACGAGCAACTACG	541	
QY	2756	TTCATATTTCGACTGAGGCGCTCACATTACCAATTATATATATCATCTCTTGGA	2815	
Db	542	TTCATTCTTCGACTGAGGTAGGACCTCACATTGATTATATACCATCTGCTTGA	601	
QY	2816	TTTG--TTCTTTATTTTGTTCAGTGATTTGATTTTCTCTGAGTACTCTCTTCCAAGTTTA	2874	
Db	602	TTTGTGTTCTTCATTTTATTTTGTGATTTGATTTGCTCTGAGTACTCTCTTCCAAGTTTA	661	
QY	2875	TTCACACAGTCGACTGTGTGATTTAAGTTTCGGGGAGGGCTCAGAA--GTATGTTGCAT	2932	
Db	662	TTCCGACAGTCGACTGTGTGATTTAAGTTTCGGGGAGGGCTCAGAAAGTGTGTGTTGCAT	721	

